

**UNIVERSITY OF NAPOLI FEDERICO II**  
**Dipartimento di Biologia e Patologia Cellulare e**  
**Molecolare “L. Califano”**

**Doctorate School in Molecular Medicine**

**Doctorate Program in**  
**Genetics and Molecular Medicine**  
**Coordinator: Prof. Lucio Nitsch**  
**XXIII Cycle**

**“Novel Transcripts from the human *DKC1* gene”**

**CANDIDATE: Dr. Alberto Angrisani**  
**MENTOR: Prof. Maria Furia**



**Napoli 2010**

# *Table of contents*

LIST OF PUBLICATIONS RELATED TO THE THESIS.....	4
ABSTRACT .....	5
1. BACKGROUND .....	6
1.1. <i>DYSKERATOSIS CONGENITA: A MULTISYSTEMIC AND GENETICALLY HETEROGENEOUS SYNDROME</i> .....	6
1.2. <i>GENES BELONGING TO THE Cbf5/DKC1/MFL FAMILY ARE INVOLVED IN A WIDE SPECTRUM OF BIOLOGICAL FUNCTIONS</i> .....	9
1.3. <i>THE STRUCTURE OF snoRNPS OF THE H/ACA CLASS</i> .....	12
2. AIMS OF THE STUDY.....	15
3. MATERIALS AND METHODS .....	16
3.1. <i>COMPUTATIONAL ANALYSIS</i> .....	16
3.2. <i>RNA EXTRACTION</i> .....	16
3.3. <i>BASIC CLONING AND MANIPULATION TECHNIQUES</i> .....	16
3.4. <i>PCR</i> .....	17
3.5. <i>qRT-PCR</i> .....	17
3.6. <i>DNA EXTRACTION FROM AGAROSE GEL</i> .....	18
3.7. <i>CLONING IN THE pGEM T-Easy VECTOR</i> .....	18
3.8. <i>CLONING IN THE p3XFLAG-CMV-10 VECTOR</i> .....	18
3.9. <i>MIDI PLASMID DNA EXTRACTION FOR CELL TRANSFECTION</i> .	18
3.10. <i>CELL CULTURE AND TREATMENT</i> .....	18
3.10.1. <i>Cell culture</i> .....	18
3.10.2. <i>Cell treatments and differentiation</i> .....	19
3.10.3. <i>Cell transfection</i> .....	19
3.11. <i>PROTEIN EXTRACTION AND QUANTIFICATION</i> .....	19
3.12. <i>WESTERN BLOT ANALYSIS</i> .....	19
3.13. <i>FLUORESCENT IMMUNOLOCALIZATION</i> .....	20
3.14. <i>HISTOLOGICAL STAININGS AND IMAGE ANALYSIS</i> .....	20
4. RESULTS AND DISCUSSION .....	21
4.1. <i>COMPUTATIONAL ANALYSIS</i> .....	21
4.2. <i>IDENTIFICATION OF DKC1 ALTERNATIVE TRANSCRIPTS</i> .....	21
4.3. <i>CODING POTENTIAL OF THE NEWLY IDENTIFIED DKC1 TRANSCRIPTS</i> .....	22
4.4. <i>QUANTITATIVE EVALUATION OF DKC1 ISOFORM EXPRESSION UNDER NMD INHIBITION</i> .....	25
4.5. <i>EXPRESSION OF DKC1 ISOFORMS DURING IN-VITRO CELL DIFFERENTIATION</i> .....	27
4.6. <i>OVEREXPRESSION OF ISOFORM 1 AND 7 IN TRANSIENTLY TRANSFECTED HeLa CELLS</i> .....	29
4.7. <i>ISOLATION AND CHARACTERIZATION OF STABLY TRANSFORMED CLONES OVER-EXPRESSING DKC1 ISOFORM 7</i> .....	31

4.8	<i>MORPHOLOGIC AND CELL-MATRIX ADHESION PROPERTIES OF CELLS OVEREXPRESSING ISOFORM 7</i> .....	32
4.9	<i>SURVIVAL OF 3XF-Iso7 TRANSFORMED CELLS AFTER X-ray TREATMENT</i> .....	38
5	CONCLUSIONS .....	41
6	AKNOWLEDGMENTS .....	41
7	REFERENCES.....	42
8	ORIGINAL PAPERS .....	46

# **LIST OF PUBLICATIONS RELATED TO THE THESIS**

Real-time PCR quantification of human DKC1 expression in colorectal cancer.  
Turano M., Angrisani A., De Rosa M., Izzo P., Furia M. *Acta Oncol.*  
2008;47:1598-9

A novel *Drosophila* antisense scaRNA with a predicted guide function.  
Tortoriello, G., Accardo, M.C., Scialò, F., Angrisani, A., Turano, M., Furia, M.  
*Gene* 2009;436(1-2):56-65

## ABSTRACT

Dyskeratosis congenita is a rare genetic disorder that causes a variety of symptoms, including mucocutaneous features, stem cell dysfunction, telomere shortening, ribosomal failure and increased susceptibility to cancer. The disease may have either autosomal dominant/recessive or X-linked inheritance. The X-linked form, that is the most common, is caused by mutations of the *DKC1* gene. Loss-of-function mutations of this gene, which encodes the nucleolar protein dyskerin, affect telomere stability, ribosome biogenesis, RNA pseudouridylation and production of small nucleolar RNA-derived micro RNAs. Dyskerin participates in fact to at least two essential ribonucleoprotein complexes: the H/ACA small nucleolarRNP complex, involved in rRNA processing and in the pseudouridylation of cellular RNAs, and the telomerase active complex, involved in telomere stability.

A single *DKC1* ubiquitary transcript has been described to date. However, the wide spectrum of functions in which the gene has been involved raised the possibility that alternative transcripts with different coding properties might also exist. Since a more extensive analysis of the *DKC1* may have shed light on the molecular mechanisms underlying the multiple functions accomplished by this gene and suggest new approaches for the therapy, I performed a detailed analysis of the gene transcriptional activity.

Here I show that at least 5 alternative transcripts, with different tissue expression profiles and coding potentials, derive from the *DKC1* gene. Intriguingly, these transcripts are differentially tuned during *in-vitro* differentiation of Caco-2 and SK-N-BE(2) cells and, although they present premature stop codons, all escape the mRNA control mediated by Nonsense Mediated Decay, which usually reduces the accumulation of aberrant transcripts.

In my thesis I also report the results obtained after constitutive overexpression in HeLa cells of one of the most interesting newly identified *DKC1* mRNAs, called Isoform 7. This isoform encodes a variant protein that is identical to canonical dyskerin but lacks the C-terminal NLS signal and is specifically -and unexpectedly- localized in the cytoplasm. Since both types of dyskerin-participated ribonucleoprotein complexes are strictly localized in the nucleus, the biological role of this variant protein is puzzling, but compatible with the hypothesis that it may be directly involved in the translation process or in transport of the small nucleolar RNA derived micro RNAs from the nucleus to the cytoplasm. Intriguingly, Isoform 7 overexpression in stable transformed cells resulted in alteration of cell morphology, increased cell-cell and cell-matrix adhesion, higher proliferation rate and increased survival after X-ray irradiation. Altogether, the results obtained indicated that *DKC1* gene expression is more complex than generally accepted, and that previously undetected alternative splice transcripts may be responsible, at least in part, for the plethora of different manifestations of the X-linked DC.

## 1. BACKGROUND

### 1.1. *DYSKERATOSIS CONGENITA: A MULTISYSTEMIC AND GENETICALLY HETEROGENEOUS SYNDROME*

Dyskeratosis congenita (DC) is a rare multi-systemic disorder characterized by heterogeneous clinical presentation and different inheritance modes. Oral leukoplakia, nail dystrophy and reticular hyperpigmentation of the skin compose a triad of mucocutaneous symptoms typical of this disease. The disease also causes premature aging, susceptibility to cancer, pulmonary diseases, osteoporosis, and bone marrow failure, with the latter representing the main cause of death. The disease strongly correlates with telomere shortening and stem cell dysfunction (for a review Kirwan and Dokal 2009).



**Figure 1. Main mucocutaneous features of Dyskeratosis Congenita.**

**Table 1. Somatic abnormalities and complications in patients with Dyskeratosis Congenita (Walne et al. 2005).**

	Abnormality	Patients, %
Classic presentation	Abnormal skin pigmentation	89
	Nail dystrophy	88
	Bone marrow failure	85,5
	Leukoplakia	78
Other common features	Epiphora	30,5
	Learning difficulties/developmental delay/mental retardation	25,4
	Pulmonary disease	20,3
	Short stature	19,5
	Extensive dental caries/loss	16,9
	Esophageal stricture	16,9
	Esophageal stricture	16,1
	Hyperhidrosis	15,3
	Malignancy	9,8

The disease exhibit genetic heterogeneity, with three kinds of inheritance modes: Autosomal Dominant (AD-DC), Autosomal Recessive (AR-DC) and X linked (X-DC), with Hoyeraal-Hreidarsson syndrome (HHS) now considered an allelic X-DC severe form.

Six causative genes have been so far identified. Out of them, three have been associated with AD-DC (MIM #127550), the milder form of the disease (Scoggins et al. 1971) characterized by generation anticipation. These genes include the telomerase RNA component gene (*TERC*; Fogarty et al. 2003), the telomerase reverse transcriptase (*TERT*; Armanios et al. 2005) and the TRF-1 interacting nuclear factor 2 (*TINF2*; Walne et al. 2008).

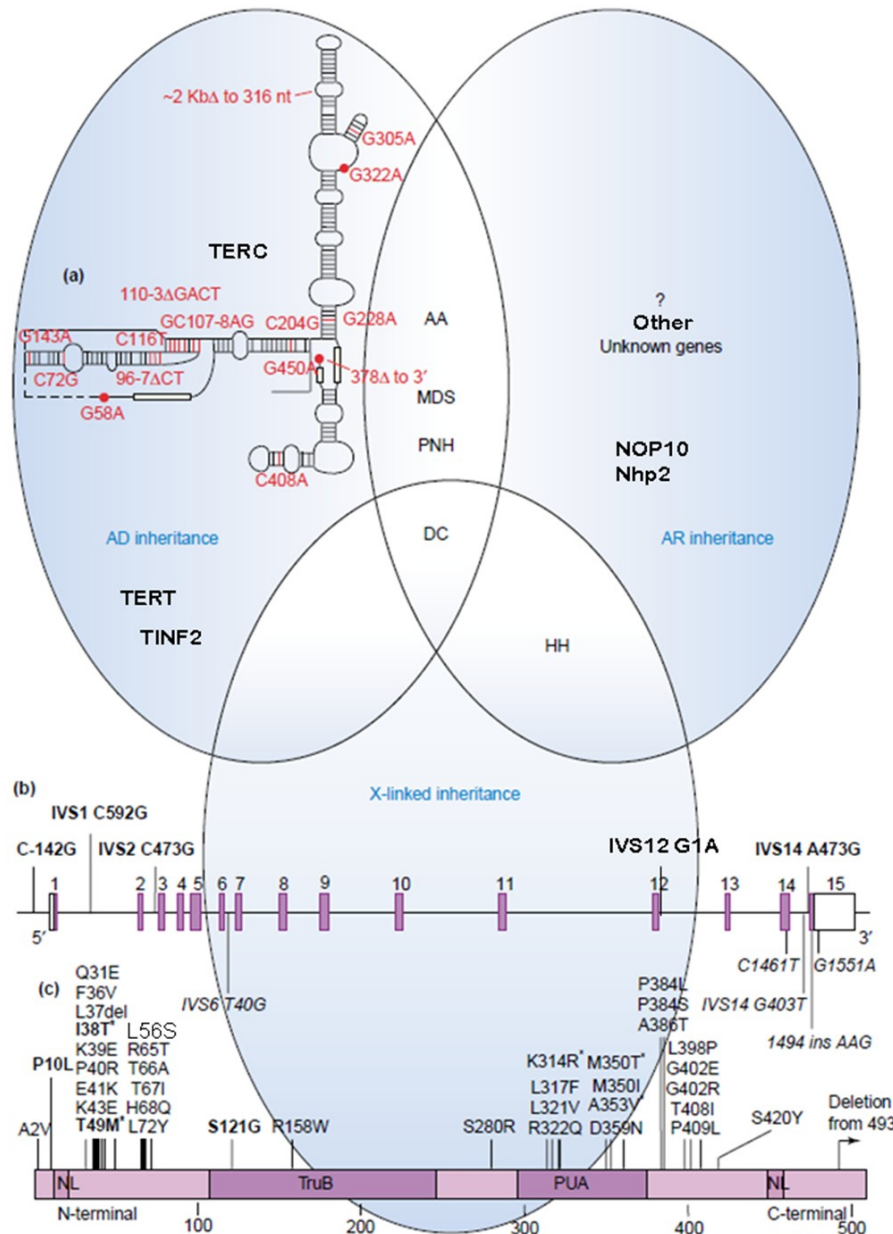
AR-DC (MIM #224230), first described by Costello and Buncke (1956), has been the most elusive form of the disease, with the *NOP10* (Walne et al. 2007) and the *NHP2* (Vulliamy et al. 2008) causative genes identified only recently.

X-DC (MIM #305000) represents the most diffuse and severe form of congenital dyskeratosis. Firstly identified by Bryan and Nixon (1965), this form is caused by mutations in the *DKC1* gene (Heiss et al. 1998), which encodes a nucleolar protein named dyskerin. Some *DKC1* mutations are also causative of HHS (MIM #300240), now recognized as a severe form of X-DC. In HHS patients, aplastic anemia, immunodeficiency, microcephaly, cerebellar hypoplasia, and fetal growth retardation are added to the classical X-DC symptoms.

More than 40 mutations affecting the *DKC1* gene have been so far identified, including an *in frame* triplet deletion (201-203 CTT) that generates the deletion of the aminoacid leucin 37 (Heiss et al. 1998); a ~2 kb (kilo bases) chromosomal deletion that causes the loss of exon 15 and the use of a cryptic polyadenilation signal (Vulliamy et al. 1999); a single mutation into the promoter region (-141; C→G) that disrupts a binding site of the transcription factor SP1, thus affecting the basal level of the promoter (Knight et al. 2001); a mutation in intron 1 (IVS1+592; C→G), that generates a cryptic donor splice site which brings to the incorporation into the mature transcript of 246 intronic bases that interrupt the CDS (CoDing Sequence) and generate a non functional transcript (Knight et al. 2001). Dyskerin mutations are mostly missense and are dispersed along the sequence, although mainly localized in the N-terminal region upstream the TruB\_N domain (TruB family pseudouridylate synthase N-terminal domain), or in 300-400 aa region, corresponding to the PUA (Pseudouridine synthase and Archaeosine transglycosylase) RNA binding domain (see Fig. 2c). Very recently, two new mutations have been identified: a two bases inversion in exon 3 (166-167invCT; Kurnikova et al. 2009), and a novel mutation within intron 12 (IVS12+1; G→A; Pearson et al. 2008). Strikingly, the latter disrupts the intron donor splice site and is predicted to lead to intron 12 retention, an event that would generate a truncated protein retaining all the functional domains with the exception of the C-terminal NLS (Nuclear/Nucleolar Localization Signal).

Dyskerin, the protein encoded by the *DKC1* gene, belongs to a family highly conserved from Archaea to mammals. Members of this family are named Cbf5p in yeast, trypanosomes and Archaea; Minifly/Nop60B in *Drosophila*, NAP57 in rat and dyskerin in mouse. In eukaryotes, proteins of this family accumulate mainly into nucleoli and are involved in several essential processes, such as ribosomal RNAs (rRNA) maturation, rRNA and

spliceosomal small nuclear RNAs (snRNA) pseudouridylation and, at least in mammals, also the maintenance of telomere integrity.



**Figure 2. Inheritance mode of DC (modified from Marrone et al. 2005).**

In a) Venn diagram summarizing DC inheritance modes. Aplastic anemia (AA), myelodysplastic syndromes (MDS), paroxysmal nocturnal haemoglobinuria (PNH) Dyskeratosis congenital (DC) or Hoyeraal-Hreidarsson (HH); in red ,mutations of the *TERC* gene. b) Mutations in *DKC1* non coding regions; in *italics*, below, known polymorphisms c) Mutations in *DKC1* coding regions; in **bold**, mutations causing HH; \* mutations recurring in more than one family.

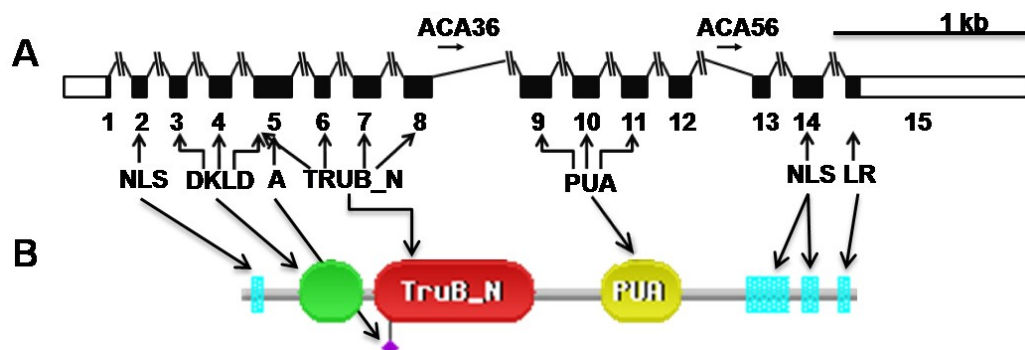
The *DKC1* gene is divided into 15 exons and hosts two snoRNAs (small nucleolar RNA) of the H/ACA class, named ACA36 and ACA56 in intron 8 and 12, respectively. Both snoRNAs have verified targets on rRNAs (see Fig.



3a). In addition, ACA56 is known to act as a micro RNA (miRNA) precursor (Ender et al. 2008).

The Pfam databank (<http://pfam.sanger.ac.uk/> accession O60832, DKC1\_HUMAN) identifies within human dyskerin three different domains. The DKLD (Dyskerin-like domain, 48-106 aa) with unknown function, but typical of this protein family; the TruB\_N the pseudouridine synthase catalytic domain, harboring the conserved aspartic 125, which marks the active site (110-226 aa), and the PUA RNA binding domain (297-370 aa). In addition, four low complexity regions (aa 11-20, 421-455, 467-480, 498-507), rich in lysine and arginine, are identified within the bipartite NLSs (see Fig. 3b). Their high level of evolutionary conservation further testifies the importance of proteins belonging to the dyskerin family. For instance, the human dyskerin and its *Drosophila* homologue, the MFL protein, share 66% of identity and 79% of similarity (Giordano et al. 1999).

Many post translational modification of human dyskerin have been annotated by the UniProt database (<http://www.uniprot.org/> accession O60832, DKC1\_HUMAN): the elimination of initiator methionine, the acetylation of alanine 2, and the phosphorylation of serines 21, 422, 451, 453, 455, 485, 494, 513 and of threonines 458, 496, 497. Altogether, these events suggest the occurrence of a complex post-translational protein regulation.



**Figure 3. Structure of the *DKC1* gene and its encoded protein.**

In A: structure of the *DKC1* gene. Only exons (boxes) are depicted in scale. Empty boxes represent 5' and 3' Untranslated Regions (UTRs); black boxes represent the CDS. ACA36 and ACA56 represent the H/ACA snoRNAs intron-encoded by the *DKC1* gene. In B: dyskerin protein structure as proposed by Pfam. In blue, low complexity regions corresponding to NLSs and Lysine Reach domains (LR); in green, DKLD domain; in red, TruB\_N domain; in purple, aspartic 125 within the catalytic site; in yellow the PUA domain.

## 1.2. GENES BELONGING TO THE *Cbf5/DKC1/MFL* FAMILY ARE INVOLVED IN A WIDE SPECTRUM OF BIOLOGICAL FUNCTIONS

In all organisms examined so far, from Archaea to Eukarya, total loss of function of members of the *Cbf5/DKC1/MFL* gene family lead to early lethality. Members of this family always proved to be indispensable for survival, and have been involved in a plethora of disparate biological functions. The first member of this family to be identified, the *Cbf5* yeast gene, was firstly described as a gene encoding a centromere binding protein (Jiang et al. 1993). Later on, gene inactivation was observed to lead to

lethality, and the lethality correlated to failure in the processing and pseudouridylation of rRNAs precursors (Zebbarjadian et al. 1999). In the same year, null mutation in the *Drosophila* ortholog gene, called *mfl*, were described to cause lethality, and *mfl* hypomorph mutations were shown to produce developmental delay, reduced fertility and inability to properly mature and pseudouridylate rRNAs precursors (Giordano et al. 1999). In trypanosomes, RNAi (RNA interference) gene silencing of the *Cbf5* gene showed that a reduction of the Cbf5p protein level led to a sharp slowdown of the growth rate, due to a drastic reduction in the synthesis of mature rRNAs. These studies also showed a general reduction in pseudouridylation levels of rRNAs, snRNAs and SLRNA (Splicing Leader RNA), which is needed for the trans-splicing mechanism (Barth et al. 2005). Controversial results on the role of members of this gene family have been gathered in mammals. In rats, the NAP57 protein has been firstly involved in "shuttling" between nucleus and cytoplasm (Meier and Blobel 1994). Subsequently, studies in primary cells obtained from patients with X-DC and from *DKC1* knock-out or mutant mice (carrying the same mutations found in human patients) unrevealed that the disease is accompanied by a reduction of telomeres length (Mitchell et al. 1999; Vulliamy et al. 2001). However, in transgenic mice the symptoms manifested many generations before the telomeres underwent a real shortening, supporting the idea that reduced levels of protein synthesis and pseudouridylation played a more important role in the pathogenesis of X-DC. When the murine *DKC1* gene was replaced by two allelic deletion variants (variant DC1, which eliminates exons 12-15, and DC7, which deletes exon 15), He and colleagues (2002) observed that females heterozygous for both deletions showed a complete embryonic lethality by day 9.5 post conception (d.p.c.). At day 7.5 d.p.c., no male embryo was vital, while female embryos showed degeneration of extra embryonic tissues. Since in these tissues the paternal X chromosome is inactivated, this suggested the occurrence of a maternal effect, so far still unexplained at the molecular level. Moreover, heterozygous females in which the deletion was of paternal origin had a strong tendency to selectively inactivate the mutant chromosome, retaining the wild type chromosome active in all cells. Since X chromosome inactivation is a very early event in embryonic development, this skewed inactivation pattern is likely to result from a "cell competition" mechanism, by which cells that inactivated the wild-type chromosome were strongly disadvantaged, and then competed-out by cells actively expressing the wild type X chromosome. Mouse embryonic cells carrying the two most common *DKC1* mutations (A353V and G402E) showed a global decrease in the levels of rRNA pseudouridylation and a reduced efficiency of pre-rRNA maturation (Ruggero et al. 2003); in addition, the A353V mutation also caused a strong destabilization of the telomerase RNA component. This feature was reflected by a significant and progressive reduction of telomere length with the proceeding of cell divisions (Mochizuki et al. 2004). RNAi *DKC1* inactivation in human cells confirmed either the defect on telomere shortening and the existence of a direct

relationship between the amount of functional dyskerin and the degree of rRNAs pseudouridylation, further supporting the dual role of this protein in these two different but essential processes (Ruggero et al. 2003; Montanaro et al. 2006).

Intriguingly, further studies on mouse cells carrying *DKC1* mutations also revealed a reduced efficiency of cap-independent mRNA translation (Yoon et al. 2006) which takes place at the IRES (Internal Ribosome Entry Site) sequences. Although active also in physiological conditions, cap-independent translation is known to have a key role in the response to apoptotic stimuli (such as  $\gamma$  radiations) and in cell cycle phases, when cap-dependent translation is reduced (i.e.  $G_0/G_1$  phases). It is thus possible that rRNA pseudouridylation may dramatically affect the ability of the ribosome to recognize IRES sequences, a fundamental step for cap-independent translation. Some antiapoptotic factors, such as XIAP and Bcl-xL, or proapoptotic, such as p27 (Kip1), are known to be translated from IRES, and their expression level was found to be reduced in *DKC1* mutant cells. These findings led to suggest that the translational imbalance between antiapoptotic and proapoptotic factors may be one of the causes of both the bone marrow failure and the tumor susceptibility observed in X-DC patients.

In conclusion, the data accumulated in mouse, rat and humans indicate that dyskerin is certainly involved in at least three basic processes: maintenance of telomere integrity; biogenesis and function of the ribosome, and pseudouridylation of various cellular RNAs. The involvement of dyskerin in maintenance of the telomere ends has been definitively established by recent experiments performed by Cohen et al. (2007), which identified by mass spectrometry three components participating to the human telomerase active complex. In addition to the two expected components, namely the telomere reverse transcriptase (*TERT*) and the telomerase RNA component (*TERC*), dyskerin was found to represent an additional, essential component of the complex. Although the association of at least 32 other proteins to the telomerase complex is suspected, none of these, except dyskerin, proved to be necessary for the catalytic activity (Cohen et al. 2007).

Intriguingly to note, novel functions have been hypothesized for the snoRNP (small nucleolar ribonucleoprotein) complexes in the last years. In fact, several reports indicated that H/ACA snoRNAs can act as precursors of miRNAs, a feature that appears to be largely widespread in eukaryotes. SnoRNA-derived miRNAs have in fact been described so far in Protozoa (Saraiya and Wang 2008), humans (Ender et al. 2008), mouse, *Arabidopsis*, yeast, chicken (Taft et al. 2009) and *Drosophila* (Jung et al. 2010). Moreover, the percentage of snoRNAs able to be processed into miRNAs has been estimated to range from 60% in humans and mouse to more than 90% in *Arabidopsis*. Although the snoRNA-miRNA processing mechanism is mostly unclear, it appears to be Dicer-1-dependent but Drosha/DCGR8-independent (Ender et al. 2008; Taft et al. 2009).

Considering that miRNAs play a fundamental role in post-transcriptional gene regulation (for a review, see Fabian et al. 2010), it is plausible that pseudouridine synthase depletion may widely affect developmental processes by inhibiting the snoRNA-derived miRNA regulatory pathway. Indeed, the occurrence of a large variety of developmental defects has recently been described in *Drosophila* as consequence of pseudouridine synthase gene silencing (Tortoriello et al. 2010). It is thus tempting to suggest that this mechanism might possibly account, at least in part, for the plethora of manifestations displayed by X-DC. Consistent with this hypothesis, dyskerin reduction has been recently shown to impair in an unpredictable way the biogenesis of snoRNA-derived miRNAs, allowing a huge spectrum of possible effects in conditions in which the expression of *DKC1* is perturbed (Alawi and Lin 2010).

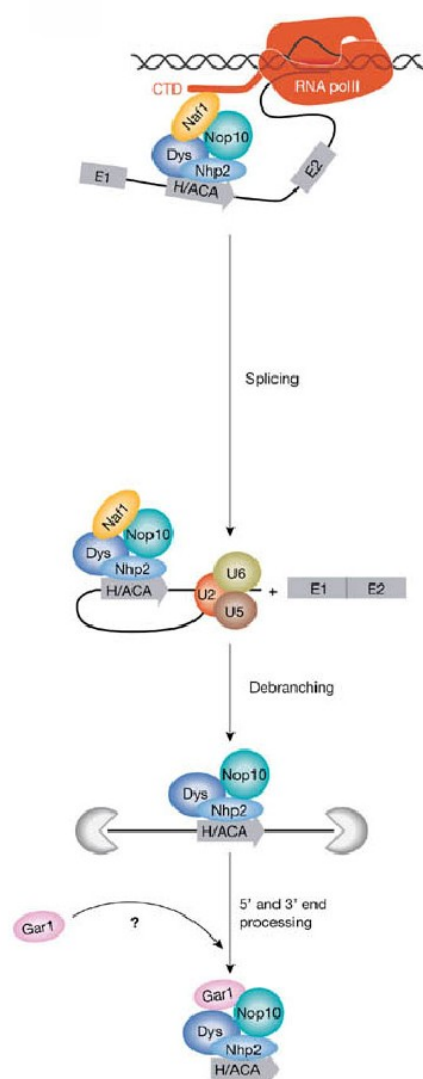
Importantly to note, the biological functions attributed to dyskerin, despite their apparent diversity, can all essentially be related to its ability to bind and stabilize snoRNAs of the H/ACA class. In fact, although the majority of these snoRNAs takes part in the pseudouridylation process, a few of them is essential for the maturation of pre-rRNA, and thus for the biogenesis of the ribosome. In addition, mammalian TERC includes a canonical H/ACA domain, thus sharing structural features with H/ACA snoRNAs (see Fig. 5). Considering the crucial importance of the snoRNP complexes, next section will be devoted to illustrate their structure and function.

### **1.3. THE STRUCTURE OF *snoRNPs* OF THE H/ACA CLASS**

Within the H/ACA snoRNP complexes, dyskerin forms a hetero-pentameric complex together with Gar1, Nop10 and Nhp2 proteins and one molecule of a snoRNA of the H/ACA class. The complex stabilizes the H/ACA snoRNAs, that otherwise would be quickly degraded.

The assembly of H/ACA snoRNPs is a very early event, since the snoRNAs, whose gene are often intronic, are bound by dyskerin before the host intron is excised by the splicing mechanism from the primary transcript (Richard et al. 2006). Intriguingly, all mammal H/ACA class snoRNAs identified till now are intron-encoded, even when their host genes do not encode proteins. Those non-coding host genes are thought to derive from progressive accumulation of mutations, with subsequent degeneration of the coding potential of the exonic sequences.

Dyskerin/snoRNA association is irreversible and established during snoRNA maturation, whereas association of Nop10, Gar1 and Nhp2 is subsequent and temporary, allowing a quick interchange between different snoRNPs (Kittur et al. 2006). Although H/ACA snoRNPs share a common proteic “core”, the catalytic activity resides on members of the dyskerin family, which catalyze the isomerization reaction. The functional specificity of the reaction is instead conferred by the particular snoRNA molecule that associates to the complex. Each snoRNA acts in fact as guide, selecting the target RNA and the specific site to be pseudouridylated by base complementarity. The typical secondary



structure of H/ACA snoRNAs is composed of two hairpins and two short single stranded sequences containing, respectively, the H (ANANNA) and ACA boxes, with the latter always localized 3 nucleotides upstream the snoRNA 3' end. Hairpin structures are interrupted by recognition loops containing the guide sequences; within the region of base pair complementarity to the target RNA, the first unpaired nucleotide represents the uridine that will be modified.

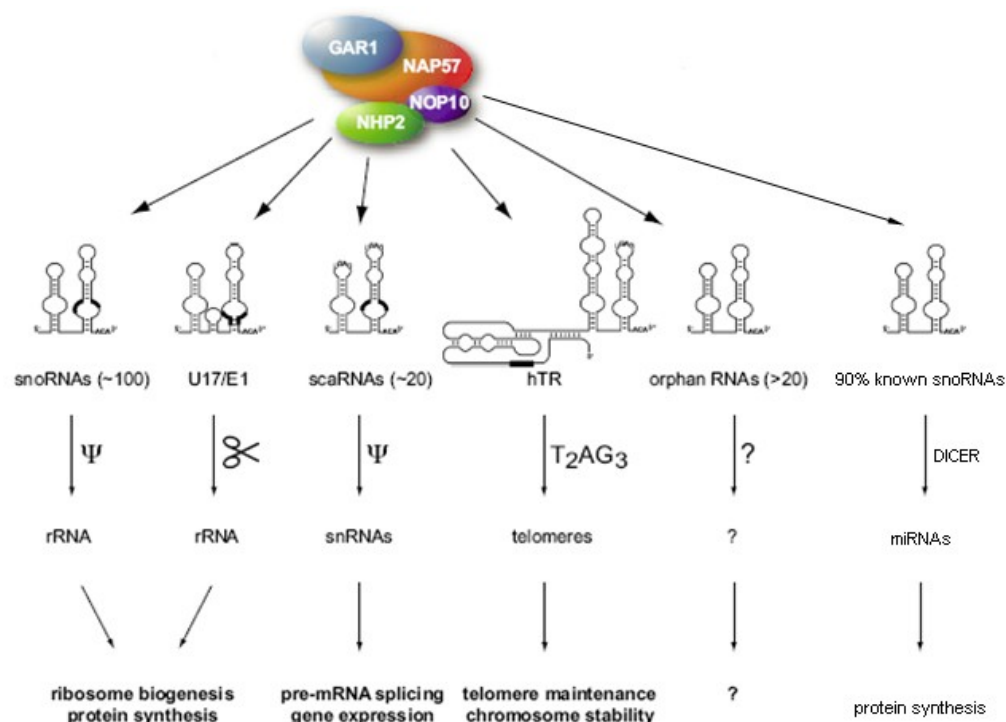
**Figure 4. Assembly of a H/ACA class snoRNP (modified from Richard and Kiss 2006).**

In mammals the assembly of H/ACA snoRNPs is co-transcriptional. The nuclear assembly factor Naf1 interacts with the C-terminal region of RNA polymerase II, aiding the binding of the H/ACA snoRNPs core proteins to intronic snoRNA sequences. The substitution of Naf1 by Gar1 determines the release of the mature particle.

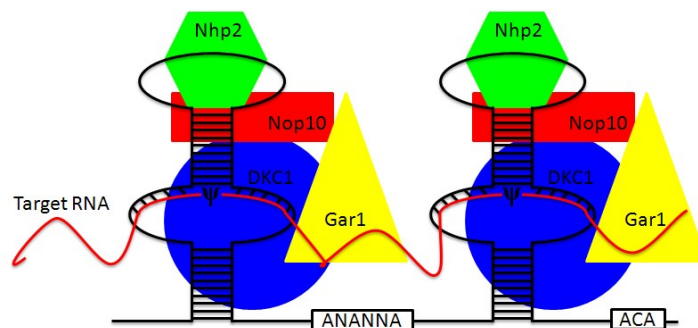
The pseudouridylation pocket is always located 14-16 nucleotides upstream of the H or the ACA box, depending on which pocket is used (see Fig. 6). The most common targets of pseudouridylation are represented by rRNAs and snRNAs, and correct pseudouridylation of these molecules seems to be necessary for proper functionality and for efficient ribosome biogenesis. However, other cellular RNAs

can be subject to pseudouridylation. In addition a growing number of snoRNAs are indicated as “orphan”, with no target RNA identified so far. Intriguingly, many “orphans” show a tissue specific expression pattern, as it is the case of HBI-36, a snoRNA specifically expressed in the brain, at the level of the choroid plexus (Cavaillé et al. 2000).

Although pseudouridylation represents the most common modification found on cellular RNAs, its biological role is still elusive. Pseudouridines are able to form hydrogen bridges with bases other than adenine, and its presence can influence folding and activity of stable RNAs, like tRNAs, rRNAs and snRNAs (Arnez and Steitz 1994; Newby and Greenbaum 2002), and also affect their interaction with RNA-binding proteins.



**Figure 5. Possible functions of H/ACA snoRNPs in mammals (modified from Meier 2005).**  
The “core” snoRNP complex, made up of 4 proteins, may bind H/ACA snoRNAs involved in different functions. The bound snoRNA determines either the target recognition and the specific biological role of the complex.



**Figure 6. Typical secondary structure of a H/ACA snoRNA molecule within the snoRNP complex.**

Some information on the structure of H/ACA snoRNP complexes has been obtained from X-ray

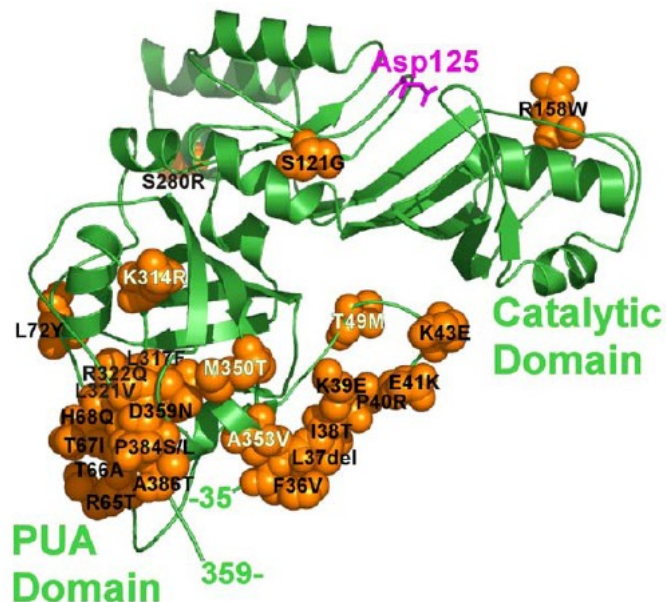
crystallography of the Archaeal *Pyrococcus furiosus* complex (Rashid et al. 2006), whose structure revealed two distinct domains, a PUA domain facing the catalytic domain, and a catalytic domain with high homology to the *E. Coli* TruB protein. The function of the PUA domain seems to be that of anchoring in the correct position the RNA helix upstream of the target uridine. Using the archaeal Cbf5p structure as template, the same authors proposed, by homology, a structural model for the structure of human dyskerin. The model, even if not complete because the archaeal protein lacks the dyskerin N-terminal (1-35 aa) and C-terminal (359-513 aa) regions, shows that most part of dyskerin pathologic mutations, even if falling far apart in the primary sequence, converge on the same side of the PUA domain. This finding suggested that these mutations may all affect the binding to substrate RNAs, or to a still



unidentified partner of the complex. Interesting to note, mutations falling in dyskerin regions missing in the archaeal protein would affect the same structural region, since those sequences are predicted to wrap the PUA domain. Finally, the role of dyskerin in ribosome biogenesis is not limited to pseudouridylation. In fact, even if by a still unclear mechanism, the protein is involved in rRNA processing under the guide of U17, E2 and E3 snoRNAs (all belonging to the H/ACA class). These snoRNAs, which guide the nucleolitic processing of rRNAs, are all essential for cell growth. In fact, maturation of 5' end of the 18S rRNA requires the activity, most probably as chaperon, of the U17 snoRNA (also called E1), which in human cells is present in two isoforms, called U17a and U17b. In turn, the snoRNA E2 is needed for the maturation of 3' end of 18S rRNA, whereas the snoRNA E3 drives the maturation of 5' end of the 5.8S rRNA. These essential functions, experimentally verified in *Xenopus laevis* oocytes (Mishra and Elicieri 1997) are evolutionarily conserved also in humans.

**Figure 7. Dyskerin structural model (Rashid et al. 2006).**

Residues mutated in X-DC patients have been superimposed to the model built on the *Pyrococcus* Cbf5p structure. Mutated residues are shown as orange spheres, mutations present in single families are printed in black, those found in more families are printed in white.



## 2. AIMS OF THE STUDY

The *DKC1* gene is causative of a congenital disease and has been involved in a wide spectrum of essential cellular processes, such as ribosome biogenesis, pseudouridylation of cellular RNAs, nucleo-cytoplasmic shuttling, maintenance of telomere integrity and snoRNA-derived miRNAs production. However, despite its essential role, the gene has been to date very poorly characterized at the molecular level. Even if *DKC1* is known to have a complex coding-non-coding organization and thus is likely to have a complex transcriptional pattern, only one 2.6 kb long constitutive transcript coding for dyskerin has been so far described and characterized (Heiss et al. 1998). Indeed, the majority of efforts have been so far focused on the identification of novel mutations and on their functional characterization in model organisms,

especially in yeast, *Drosophila* and mice. On the other hand, considering that one of the steps needed for cancer progression is the reactivation of telomerase function, additional approaches have been addressed to the comprehension of the role played by *DKC1* activity in tumorigenesis. In fact, the possibility to correlate the levels of *DKC1* expression to cancer progression is particularly attractive and may possibly furnish an easy and early marker of the process. To this regard, I contributed to show that up-regulation of *DKC1* gene during colon cancer development may represent a more sensitive index than up-regulation of telomerase activity (Turano et al. 2008).

In order to increase our knowledge on the molecular mechanisms underlying the pathogenesis of X-DC, during my PhD thesis I decided to perform a detailed analysis of *DKC1* transcriptional activity. I undertook this approach because I thought very plausible that the wide spectrum of cellular functions performed by the *DKC1* gene may, at least in part, rely on the production of alternative transcripts.

Alternative splicing is widely used by eukaryotic genes to generate multiple transcripts and proteic isoforms often playing different biological roles, sometimes synergic and sometimes antagonistic. Alternative splicing can in fact dramatically change the structure of the mature transcripts and their coding potential. For many alternative transcripts, even if characterized by a low-level, or a transient expression profile, a fundamental role in many cellular processes has been established. Taking into account the NMD (Nonsense Mediated Decay) mechanism of mRNA surveillance, it has been estimated that 25% of alternative exons is fundamental to determine transcript abundance (for a review, Stamm et al. 2005). Finally, all the bulk of data present in the recent literature indicates that alternative splicing can regulate gene expression in a quicker and finer way than the regulation of promoter activity. Indeed, the results I have obtained confirmed that *DKC1* gene expression is more complex than generally accepted, and support the possibility that previously undetected alternative transcripts may be responsible, at least in part, for the plethora of different manifestations of the X-linked DC.

### **3. MATERIALS AND METHODS**

#### **3.1. COMPUTATIONAL ANALYSIS**

Nucleotide and aminoacid reference sequences were those present at the GeneBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Translate and ProtParam analysis tools at the ExpASy web site (<http://www.expasy.ch/>) were used.

#### **3.2. RNA EXTRACTION**

Total RNAs were extracted from tissues and cultured cells by the TRI-reagent (Sigma) following manufacturer's instructions; RNAs were dissolved in a variable amount of 0.1% DEPC treated water. Samples' were quantified by the Nanodrop instrument (Thermo scientific).

#### **3.3. BASIC CLONING AND MANIPULATION TECHNIQUES**



Basic cloning techniques, nucleic acids and proteins manipulations, SDS-PAGE and wet protein blotting were carried out according to Sambrook and Russell (2001).

### 3.4. PCR

Qualitative PCRs were performed using High Fidelity PCR Master Mix (Roche) using manufacturer's instructions and carried out by the MyCycler thermal cycler (BioRad). All of the primers were designed by the Primer3 program (<http://frodo.wi.mit.edu/>) and have an annealing temperature of circa 60°C.

The amplification protocol was:

94°C 2 minutes for enzyme activation and cDNA denaturation

10 cycles composed of:

Denaturation: 94°C 10 seconds

Annealing: 60°C 70 seconds

Polymerization: 72°C, the time depended on fragment length, considering a 1 kb per minute rate

20 cycles composed of:

Denaturation: 94°C 15 seconds

Annealing: 60°C 30 seconds

Polymerization: 72°C, the time depended on fragment length, considering a 1 kb per minute rate and an increase of 5 seconds per cycle

Then a single final extension: 72°C 7 minutes

The reaction products were verified on agarose gel stained by ethidium bromide.

Oligonucleotides used for cloning the isoforms are:

ISOFORM	FORWARD PRIMER	REVERSE PRIMER
Commons	AGGGTAACATGGCGGATG	TCAACAAGAAAACCCAACAGG
Isoform 3	TGTGTTTGACTTCACTTTGACTAAA	GGTAAGGCCAGCACCATTTA
Isoform 4	GGAGGGAAGGACTGGCTAGA	CCAGCAGACAAAACATAGAAGG
Isoform 5	ATCTCCTGACGTCGTGATCC	GTAGAGACGGGGTTTCACCA
Isoform 6	TGGGCTAAGGGAGATGATTG	CTCCCAAAGTGCTGGGATTA
Isoform 7	ATGGGAAGGTAGAGCCCACT	CTGAAACATCCTGCCCTCAC

### 3.5. qRT-PCR

The Real time PCR samples preparation, amplification and analysis were performed as described elsewhere (Turano et al. 2008). The amounts of *DKC1* gene transcripts in different samples were normalized respect to housekeeping gene. The housekeeping genes used were: GAPD for tissues comparison, POLR2A for Caco-2 and SK-N-BE(2) differentiation and GSS for NDM inhibition, the sequences of the primers used are:

GENE	FORWARD PRIMER	REVERSE PRIMER
POLR2A	TGCTCCGTATTGCGCATCATG	TCCATCTTGTCACCACTCTT
GAPD	ACAGAGTGAGCCCTTCTTCAA	GGAGGCTGCATCATCGTACT
GSS	GGACTGGCCCTGGGAATT	CCTTCTCTTGAGCAATCAGTAGCA

The sequences of primers used for *DKC1* transcripts amplification are:

ISOFORM	FORWARD PRIMER	REVERSE PRIMER
---------	----------------	----------------

Isoform 1	CTCGGAAGTGGGGTTTAGGT	ACCACTTCAGCAACCACCTC
Isoform 3	GGCGGATGCGGAAGTAAT	TTTAGTCAAAGTGAAGTCAAACACA
Isoform 4	TTTTCTGCTGTGCAAAAT	GCTGGAATAAGGCACCTGTC
Isoform 5	GGCTGCACAATGCTATTGAA	CATACCCACTCACAGGCCTAA
Isoform 6	CTCCAGCCTACAAAAATGC	CACATCATGCATTGTCACCA
Isoform 7	CTCGGAAGTGGGGTTTAGGT	CTGAAACATCCTGCCCTCAC

The sequences of primers used as treatments controls are:

GENE	FORWARD PRIMER	REVERSE PRIMER
rpL3	ATGAATGCAAGAGGCGTTTC	CTAAGGAGCCTTTTCCACCA
S100G	CCAGACCAGTTGTCAAAGGA	TCCATCTCCATTCTTGTCCTCA
RET	CCACGGTGGCCGTGAA	CTGACAGCAGGTCTCGAAGCT

All oligo pairs produce an amplicon of circa 150 bp.

All the primer pairs employed are separated by at least one intron so to prevent the amplification of contaminating genomic DNA.

### 3.6. ***DNA EXTRACTION FROM AGAROSE GEL***

PCR products were loaded on agarose gel and run for 45 minutes at 100V to ensure fragments separation. The bands of interest were cut and DNA fragments eluted by the Nucleospin Extract II kit (Macherey-Nagel) following manufacturer's instructions.

### 3.7. ***CLONING IN THE pGEM T-Easy VECTOR***

The plasmid used for sequencing reactions is the pGEM T-Easy vector (Promega), ligation and transformation were performed following manufacturer's instructions.

### 3.8. ***CLONING IN THE p3XFLAG-CMV-10 VECTOR***

The fragments corresponding to Isoform 1 and 7 CDS were amplified using 3 different primers. These primers contain, upstream to the sequence complementary to transcripts, sequences target of specific restriction enzymes for directional cloning in the expression vector. The primers are:

5'CCCAAGCTTGGTAACATGGCGGATGCG3' located in exon 1, contains the AUG, the HindIII site is underlined, is common to both fragments

5'CCGGAATTCTCCAGCTTCAAGTGGCC3' located in exon 15, the EcoRI site is underlined, defines the stop codon of Isoform 1

5'CCGGAATTCACCCCGGCTCTGAAACATC3' located in intron 12, the EcoRI site is underlined, defines the stop codon of Isoform 7

The PCR products were first purified, then subject to double enzymatic restriction. Fragments were again purified and ligated into the p3XFLAG-CMV-10 plasmid using a 6:1 (insert/plasmid) ratio.

### 3.9. ***MIDI PLASMID DNA EXTRACTION FOR CELL TRANSFECTION***

To obtain endotoxins free plasmid positive clones were extracted by the GeneElute™ HP Plasmid Midiprep kit (Sigma) following manufacturer's instruction, the plasmids were stored at +4°C.

### 3.10. ***CELL COLTURE AND TREATMENT***

#### 3.10.1. ***Cell culture***

The cell lines used were: MCF-7 breast adenocarcinoma, HeLa cervix adenocarcinoma, ZR-75.1 breast ductal carcinoma, Caco-2 colorectal adenocarcinoma and SK-N-BE(2) neuroblastoma. Cell lines were grown in DMEM (Dulbecco's Modified Eagles Medium) supplemented with FBS 10% (MCF-7, HeLa and ZR-75-1), 15% (SK-N-BE(2)) or 20% (Caco-2); 2 mM L-glutamine, penicillin 100 U/ml and streptomycin 100 µg/ml at 37°C 5% CO<sub>2</sub> in a humidified chamber, all chemicals were from Invitrogen.

### **3.10.2. Cell treatments and differentiation**

To evaluate NMD regulation, protein translation of Caco-2 cells was blocked by cycloheximide 100 µg/ml for 2 hours. Caco-2 cells were induced to differentiate as described by Pinto et al. (1983) and Pignata et al. (1994). To induce the differentiation of SK-N-BE(2), cells were treated as described elsewhere (Tahira et al. 1991; D'Alessio et al. 1995; Corvi et al. 1997). Growth curve were calculated counting cells by a Burker chamber and seeded  $5 \times 10^5$  cells per 5 cm culture dish. They were counted for successive 4 days. The experiment was repeated 3 times. Cell-matrix adhesion was evaluated as described by Wang and colleagues (2010). For X-ray surviving assay cells were seeded on plastic culture dishes and when confluence was at 50% cells were subjected to increasing doses of X-ray: 0, 2, 4, and 6 Gy at a rate of 1 Gy/minute. Then cells were detached by trypsin, counted and seeded at serial dilution to evaluate cloning potential. After 8 days clone density was near the ideal 1 clone every 50-200 cm<sup>2</sup> of culture dish so cells were fixed with 10% methanol 10% acetic acid for 10 min at room temperature, washed with PBS and stained with Crystal Violet solution. The experiment was repeated trice.

### **3.10.3. Cell transfection**

For transient and stable transfection HeLa cells and Lipofectamine reagent (Invitrogen) according to manufacturer's instructions were used. For transient transfection  $10^6$  cells were seeded into a 10 cm culture dish containing 1 microscope slide coverslip for fluorescent immunolocalization and transfected by 8 µg of each plasmid. For stable transfection  $2 \times 10^5$  cells were seeded into 5 cm culture dishes and transfected by 1 µg of each plasmid, stably transfected clones were selected by G418.

## **3.11. PROTEIN EXTRACTION AND QUANTIFICATION**

For protein extraction, cells were washed twice with ice cold PBS, detached from dish by a cell scraper and harvested in 4 ml of PBS. Cells then were pelleted by centrifugation for 5 minutes 1400xg at 4°C. The PBS was eliminated and cells resuspended by gentle vortex in 100 µl of RIPA lysis buffer. After 30 minutes of incubation in ice, the cell lysates were transferred in micro centrifuge tubes and centrifuged for 15 minutes 14000 rpm at 4°C to eliminate insoluble material. The supernatants were transferred in new tubes and stored at -80°C. Protein concentration of extracts was determined by the BioRad reagent following manufacturer's instructions.

## **3.12. WESTERN BLOT ANALYSIS**

The membranes produced by proteins blotting were washed three times for 5 minutes with TBS-t-m (TBS-t containing 5% non fat dried milk). Then membranes were incubated with anti-FLAG antibody (Sigma) diluted 1:5000 in TBS-t-m for 2 hours at RT. The membranes were washed three times for 5 minutes with TBS-t; then they were exposed to anti-mouse antibody (SantaCruz) conjugated with horseradish peroxidase diluted 1:5000 in TBS-t-m. Membranes were washed again in same conditions and exposed to ECL-Plus (GE Healthcare) for 5 minutes. The excess of ECL-Plus was eliminated and chemiluminescence was visualized exposing the membrane into the ChemiDoc (BioRad) and images were saved on a PC.

### **3.13. *FLUORESCENT IMMUNOLocalIZATION***

Fluorescent immunolocalization was performed by a secondary antibody conjugated to TexasRed. Cells attached to coverslips were washed for 5 minutes with PBS (thereafter simply washed), then fixed for 10 minutes at RT with 3.7% formaldehyde in PBS. Samples were washed then cells were permeabilized by 0.5% Triton X-100 in PBS for 15 minutes at RT then washed. Primary anti-FLAG antibody diluted 1:250 in PBS with 0,02% Tween was added and incubated for 1 hour. From this point the steps were performed in dark; samples were washed and secondary anti-mouse antibody diluted 1:250 in PBS with 0,02% Tween was added and incubated 1 hour. Coverslips were washed and exposed to DAPI diluted 1:5000 in PBS for 5 minutes to counterstain nuclei after a final wash slides were mounted by PBS with 50% glycerol and coverslip borders sealed with nail paint

### **3.14. *HISTOLOGICAL STAININGS AND IMAGE ANALYSIS***

Cell blocks were prepared harvesting cells in Versene solution (PBS plus 53 mM EDTA). Cells were centrifuged for 6 minutes at 4000rpm then fixed by 10% formalin in ethanol for 45 minutes; alternatively cells were grown on coverslips and directly fixed by 3.7% formalin in PBS and stained. Fixation of cells grown on coverslips was performed like in immunofluorescence. Standard haematoxylin eosin, Papanicolaou, silver nitrate active NORs stainings were performed according to Bancroft and Gamble (2007).

Deparaffinization and rehydration:

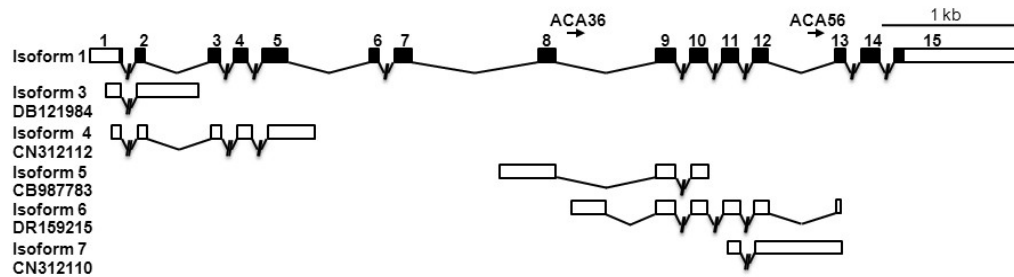
- Xylene 10 minutes
- Ethanol 100% 10 minutes
- Ethanol 80% 10 minutes
- Ethanol 70% 10 minutes
- Ethanol 50% 10 minutes
- Distilled water 10 minutes

Final dehydration and clearing were performed with same times and concentrations but in opposite order. Haematoxylin eosin stained sections were photographed under light transmission microscope, nuclear and cytoplasmic areas were measured on acquired images by the ImageJ software (<http://rsbweb.nih.gov/ij/>); mean, standard deviation and one tailed Student's t-test were calculated by Excel software (office.microsoft.com).

## 4. RESULTS AND DISCUSSION

### 4.1. COMPUTATIONAL ANALYSIS

Starting from *DKC1* (GeneID: 1736) known sequences (NC\_000023 for gene, NM\_001363 for the 2.6 kb ubiquitary Isoform 1 and NM\_001142463 for the predicted Isoform 2), I searched the dbEST database by the BLAST algorithm (Basic Logic Alignment Search Tool), using as queries the sequences of all introns of the gene. The GenBank dbEST represents a powerful tool for analysis of gene expression and the search of alternative transcripts. My search resulted in the identification of 5 *DKC1* ESTs (Expressed Sequence Tags) with accession numbers: DB121984, CN312112, CB987783, DR159215, CN312110, here referred as Isoform 3, 4, 5, 6 and 7 respectively. Alignment of each EST on the genomic sequence of the *DKC1* gene is reported in Fig. 8. As shown in the figure, all the identified ESTs were characterized by partial or complete retention different *DKC1* gene introns.



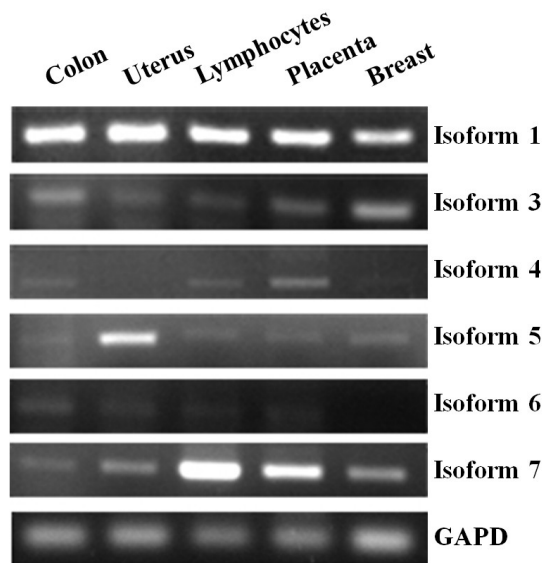
**Figure 8. Structure of the *DKC1* gene and *DKC1*-derived ESTs annotated on the dbEST database and characterized by intron retention.**

On the top, structure of the *DKC1* 2.6 kb canonical transcript (Isoform 1). Only exons (boxes) are depicted in scale. Empty boxes represent 5' and 3' UTRs, black boxes represent CDS, ACA36 and ACA56 represent the H/ACA snoRNAs intron-encoded by the gene. Below, the structures of the five ESTs identified in this study.

### 4.2. IDENTIFICATION OF *DKC1* ALTERNATIVE TRANSCRIPTS

Starting from the structure of the five ESTs identified, I attempted to confirm their expression and define the complete structure of the putative transcripts from which they should derive. With this aim, total RNA was extracted from two breast cancer cell lines (MCF-7 and ZR-75.1) and from five normal tissues (breast, colon, uterus, placenta and lymphocytes) and RT-PCR experiments focused to specifically detect the expression of EST-related transcripts were performed. The results obtained confirmed the effective accumulation of the predicted transcripts in either the examined cell lines (not shown) and the tested tissues (see Fig. 9). These analyses also revealed that the levels of expression of these transcripts could be differentially modulated in the diverse tissues. In fact, while Isoform 1 was found expressed at similar levels in all tissues, Isoform 4 was undetectable in the uterus, while Isoform 6 in breast. In contrast, Isoform 5 proved to be up-regulated in uterus, while Isoform 7 was up-regulated in lymphocytes and placenta (Fig. 9). Even if these preliminary data were only semi-quantitative, they confirmed the actual expression of previously undetected *DKC1* alternative transcripts. However,

neither in normal tissues or cell lines we were able to amplify any fragment related to Isoform 2 expression. It seems possible that this dbEST-annotated isoform might actually derive from an artifact, or alternatively, from a transcript specifically expressed in a tissue not examined in our experiments. I cannot in fact exclude that expression of the Isoform 2-related transcript can be restricted to the embryonal carcinoma cell line, from which the only EST representative of this hypothetical transcript, BG258928, has been annotated.



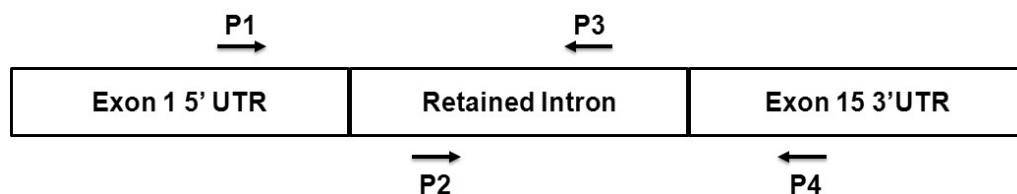
**Figure 9. Expression profiles of alternative *DKC1* transcripts determined by RT-PCR experiments.**

The constitutive and ubiquitous Isoform 1 is expressed at a similar level in all the tested tissues, whereas the expression of newly identified isoforms appears to be differently modulated in the diverse cell types. Expression of the ubiquitously and constitutively expressed GAPD (Glyceraldehyde-3-phosphate dehydrogenase) gene is shown for internal normalization.

In order to define the complete structure of the newly identified transcripts, I performed additional RT-PCR experiments focused to clone full length cDNAs using as

template RNA extracted from the MCF-7 cell line. The strategy followed in these experiments (depicted in Fig. 10) was aimed to amplify the regions upstream and downstream each EST annotated sequence, and made use of primers derived from the first and the last *DKC1* exons (exon 1 and exon 15), coupled with isoform-specific intronic primers.

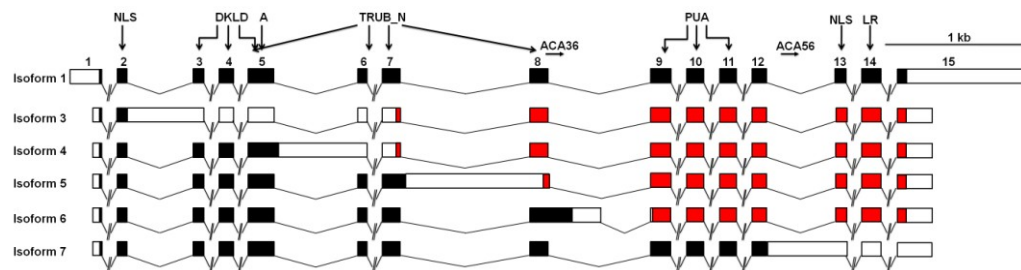
PCR products were loaded on agarose gel to check the correct size and then cloned into the pGEM T-Easy vector for sequencing reactions. Positive clones were sent to the external company Primm (Primm, Milan Italy) for nucleotide sequencing.



**Figure 10. RT-PCR strategy followed for the amplification of the novel *DKC1* transcripts.** Primers P1 and P4 were common to all the isoforms and were located outside of the CDS contained within the canonical Isoform 1, whereas primers P2 and P3 were isoform-specific and derived from the specifically retained intron.

#### **4.3. CODING POTENTIAL OF THE NEWLY IDENTIFIED *DKC1* TRANSCRIPTS**

Nucleotide sequence analysis revealed that *DKC1* alternative transcripts differed from Isoform 1 only for the retention of specific introns. When the coding potential of the new isoforms was evaluated by the online program “Translate” of the ExPASy web site, a software that enables the simultaneous search of CDS in all the three possible frames, I observed that the CDS starting at the dyskerin first AUG codon was maintained in all of them. However, with the exception of Isoform 7, these CDS were significantly shorter than that of canonical dyskerin, because of the occurrence of premature stop codons (PTCs) at different points. As consequence, the putatively encoded proteins would miss one or more carboxy-terminal dyskerin functional domains (see Fig. 11). Strikingly, these upstream CDSs were followed by significantly longer, and *in frame*, CDSs which, conversely, would miss N-terminal or internal dyskerin functional domains.



**Figure 11. Complete structure and coding potential of the new *DKC1* transcripts.**

Nucleotide sequencing of the novel isoforms demonstrated that dyskerin CDS (black boxes) is precociously interrupted in most of them, thus being able to direct the synthesis of truncated proteins missing one or more C-terminal functional domains. Note presence of longer, downstream CDSs (red boxes) within isoforms 3,4,5, and 6.

The coding properties of novel *DKC1* isoforms are described here in more detail. Isoform 3, that is characterized by retention of the whole intron 2, includes a very short upstream CDS that is able to encode a peptide, only 28 aa long, of unpredictable function.

Isoform 4 is instead characterized by retention of the whole intron 5, and its upstream CDS potentially encodes a protein of 159 aa (18 kDa) which would include the N-terminal NLS, the DKLD domain and part of the TruB\_N domain, comprising the catalytic aspartic residue. This short protein would lack the PUA RNA binding domain, leading to exclude that it can act as catalytic pseudouridine synthase.

Isoform 5 upstream CDS retains intron 7, and would encode a protein of 219 aa (27.7 kDa) which includes the N-terminal NLS, the DKLD and most of the TruB\_N domain, including the catalytic aspartic residue. However, also this truncated protein lacks the PUA RNA binding domain, and thus should be unable to act as catalytic pseudouridine synthase.

Isoform 6 upstream CDS is characterized by partial retention of intron 8 and would encode a protein of 313 aa (35.1 kDa) which includes the N-terminal NLS, the complete DKLD and TruB\_N domains but still lacks the PUA domain, which is essential for RNA binding. Worth noting, this isoform derives from the usage of an internal donor splice site located downstream the

intron-encoded snoRNA ACA36, coupled with the canonical downstream acceptor splice site and polypyrimidine tract. As consequence of this splicing pattern, the snoRNA ACA36 is exonized, and this event might possibly antagonize its biogenesis through the splicing-dependent mechanism.

As mentioned above, each of the above listed isoforms present an additional downstream CDS, starting at an internal *in frame* AUG codon. These CDSs are significantly longer than the upstream, and all potentially encode variant proteins containing the C-terminal dyskerin moiety. For example, for both isoforms 3 and 4, a downstream CDS starting at an AUG codon located in exon 7 would encode a protein of 311 aa with a weight of 34.8 kDa (kilo Dalton), which is predicted to contain the C-terminal region of the TruB\_N domain, devoid of the catalytic aspartic, the PUA domain and the C-terminal NLS.

Similarly, Isoform 5 presents a downstream CDS starting at an AUG codon located in exon 8. This CDS would encode a protein of 272 aa (30.4 kDa) composed of the C-terminal moiety of the TruB\_N domain, which included the catalytic aspartic, the PUA domain and the C-terminal NLS.

Finally, within Isoform 6 an internal AUG located at exon 9 marks a CDS able to produce a protein of 255 aa (28.4 kDa), which includes only the PUA domain and the C-terminal NLS.

Although synthesis of these putative proteins would depend on an internal IRES, whose activity remains to be demonstrated, it is interesting to note that these products are all predicted to retain the ability to bind RNA, but would lack the catalytic pseudouridine synthase activity. This feature is compatible with the possibility that, if produced, they might act as dyskerin dominant-negative regulators. Alternatively, it is plausible to hypothesize that they can simply act as snoRNA-chaperones, with no active role in the pseudouridylation process.

Noticeably, Isoform 7 is the only transcript that, similarly to Isoform1, contains a single, long CDS. This isoform differs from Isoform 1 only for retention of intron 12, an event that would lead to the exonization of the snoRNA ACA56. As mentioned above, this event might antagonize the snoRNA biogenesis through the splicing-dependent mechanism and might have functional relevance, since this snoRNA is known to act as miRNA precursor (Ender et al. 2008). Moreover, Isoform 7 particularly attracted my interest because it would encode a 420 aa protein (47.6 kDa) that includes all the dyskerin functional domains, with the exception of the C-terminal NLS. This putative protein was likely to keep the pseudouridine synthase catalytic activity, although its nuclear localization could possibly be affected by the absence of the C-terminal NLS.

Expression of this protein might be of functional relevance also in the light that Pearson et al. (2008) reported a case of a HHS male child patient carrying a *DKC1* mutation at position +1 of intron 12. This mutation would lead to intron 12 retention, exactly as observed for Isoform 7, and was not embryonic lethal, although the child patient died at age 2 years. Altogether, these observation made Isoform 7 very attractive for further studies.



To get a more complete information about all the *DKC1* putatively encoded variant proteins, I also analyzed their sequences with the “ProtParam tool” at the ExPASy web site for a detailed prediction of their weight, isoelectric point (pI) and instability index (II). This latter parameter, based on the composition of the first dipeptide of the protein, was calculated on all the 400 possible dipeptides by Guruprasad and colleagues (1990). Proteins with II values of 40, or below, are predicted to be stable, whereas those with higher values are presumably unstable.

**Table 2: Predicted parameters of *DKC1* putative alternative isoforms.**

Isoform	Frame 1 aa	kDa	pI	II	S/U	Frame 2 Aa	kDa	pI	II	S/U
1	513*	57.5	9.46	44.77	U	N/A	N/A	N/A	N/A	N/A
2	508 <sup>#</sup>	56.9	9.50	45.03	U	N/A	N/A	N/A	N/A	N/A
3	28	3.2	9.22	64.84	U	311	34.8	9.25	45.00	U
4	159	18.0	9.40	42.36	U	311	34.8	9.25	45.00	U
5	219	27.7	9.55	49.28	U	272	30.4	9.39	44.66	U
6	313	35.1	9.51	46.24	U	255	28.4	9.35	39.31	S
7	419 <sup>#</sup>	47.5	9.40	44.01	U	N/A	N/A	N/A	N/A	N/A

Frame 1 aa: always starting from the Isoform 1 canonical AUG; Frame 2 aa: starting from a downstream AUG, located in exon 7 for Isoform 3 and 4, in exon 8 for Isoform 5 and in exon 9 for Isoform 6; aa: length in aminoacids, kDa: weight in kilo Dalton, pI: isoelectric point, II: instability index score, S/U: result of II, S: stable, U: unstable. \* the analysis was performed on the mature form, missing of the initial methionine; # it was supposed an elimination of initial methionine similar to Isoform 1; N/A not allowed.

The properties of all putatively encoded proteins are listed in Table 2. *DKC1* variant proteins are all predicted to have a quite basic pI (ranging from 9.22 to 9.55), most probably for the high content of lysines and arginines at the terminal NLSs. Intriguingly, on the basis of their II, all are predicted to be quite unstable, with the exception of the product of Isoform 6 downstream CDS. Strikingly, also the canonical dyskerin (Isoform 1) is predicted to be quite unstable. Even if it is not possible to judge the effect of the acetylation of alanine 2 on the predicted II, this observation unravels that this fundamental protein might have a fast turnover, so far unnoticed. Finally, it is worth noting that most of the *DKC1* variant isoforms are predicted to differ from canonical dyskerin also for several post-translational modifications, a feature that may also be of wide functional relevance.

#### **4.4. QUANTITATIVE EVALUATION OF *DKC1* ISOFORM EXPRESSION UNDER NMD INHIBITION**

The production of a translationally competent mRNA is the result of the combination of transcription, maturation, nuclear export and maturation of the messenger ribonucleoparticle. In higher eukaryotes, the quality control of mRNAs is a fundamental process. In fact, the elimination of anomalous transcripts that contain PTCs, and are thus able to drive the synthesis of truncated proteins with detrimental effects (like, for example, negative dominance), is of primary importance. One of the mechanisms that performs this control is named NMD (for a review Maquat 2004; Amrani et al. 2006). PTC-containing transcripts may be generated in different ways: from

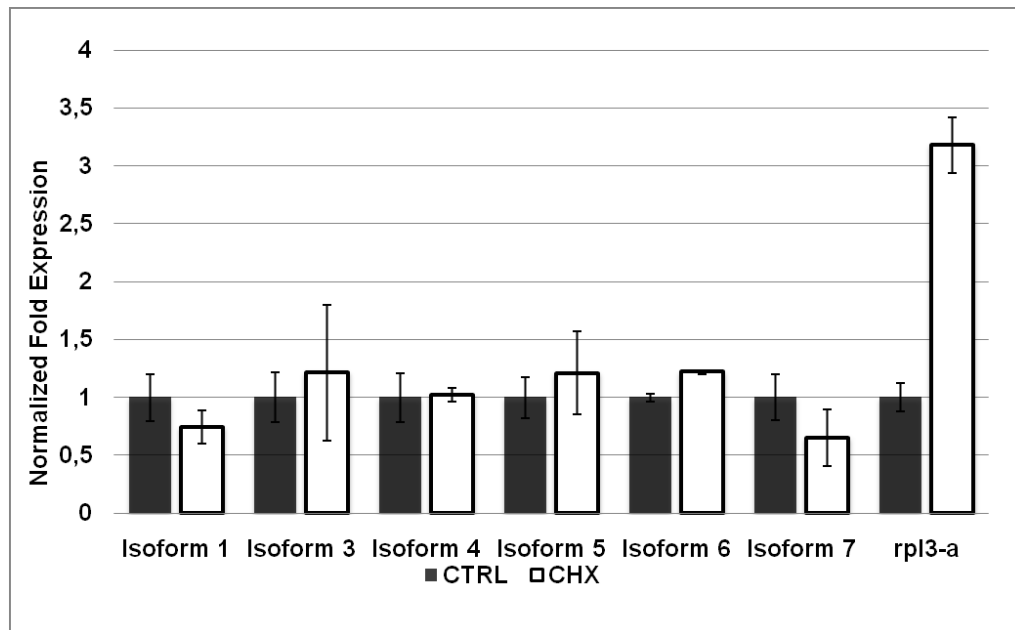
transcription of a mutated gene, from errors during transcription, from incorrect splicing that generates frame-shifting, from missing or incomplete intron elimination, from transcription of a pseudogene, etc. However, production of PTC-containing alternative transcripts may also represent an important physiological mechanism of self-regulation of gene expression, aimed at down-regulate the level of the encoded protein (Lewis et al. 2003). Since all the newly identified *DKC1* transcripts were characterized by presence of PTCs, I wished to check whether their expression could be down-regulated by the NMD surveillance mechanism, as expected for physiological aberrant transcripts. Position of PTCs within isoforms 3, 4, 5, 6 and 7 respect the downstream exon-exon junction is summarized in Table 3.

**Table 3. Position of PTCs within *DKC1* alternative isoforms.**

Isoform	nt*	EJ
3	542	3-4
4	629	6-7
5	1030	8-9
6	199	8-9 <sup>§</sup>
7	564	13-14

\* nucleotides upstream of exons junction (EJ); § referred to the internal splice site of intron 8.

To be activated, NMD needs ongoing protein synthesis, so the simplest way to highlight the occurrence of this regulatory mechanism is to inhibit protein synthesis by culturing cells in a medium containing the appropriated antibiotics. I then treated Caco-2 cells with cycloheximide and analyzed by reverse transcription quantitative Real time PCR (qRT-PCR) the total RNA extracted from both treated and untreated control cells after a set-up phase, necessary to check the efficiency of the primer pairs to be used in transcript quantification. In qRT-PCR experiments, the expression of the glutathione synthetase gene was used to normalize the RNA amount among different samples, while the expression of an alternatively spliced form of the *RPL3* gene, that is known to be under the control of NMD (Cuccurese et al. 2005), was employed as positive control. As expected, under NMD inhibition the expression of Isoform 1 was unaffected. Surprisingly, the expression of alternative *DKC1* transcripts, all containing PTCs, was similarly unaffected (Fig. 12). This finding suggests the possibility that these transcripts may have functional significance, and poses intriguing questions about their possible functions and on the molecular mechanisms by which they escape the NMD surveillance.

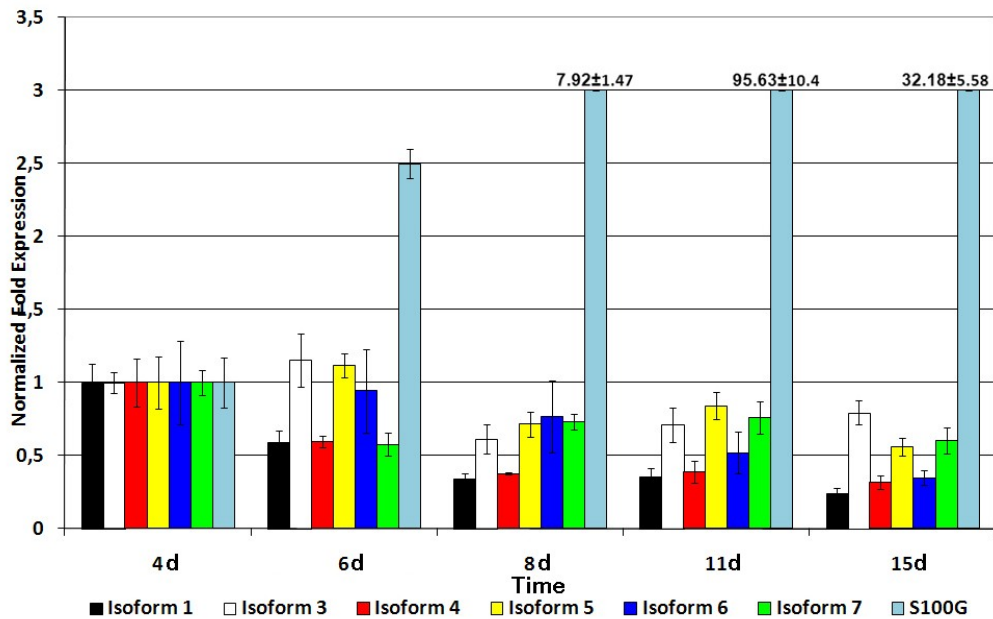


**Figure 12. Expression of different isoforms produced by the *DKC1* gene in Caco-2 cells treated with cycloheximide.**

The levels of all isoforms were unaffected by NMD inhibition. CHX: cycloheximide-treated cells; CTRL: control untreated cells. The expression of the rpl3-a alternative transcript is shown as positive control of NMD inhibition.

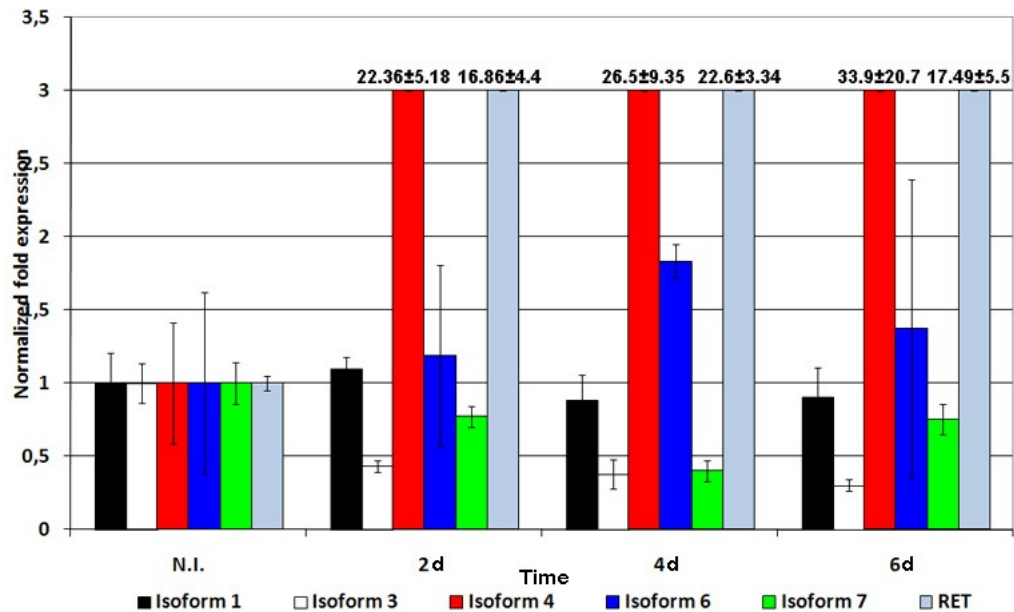
#### 4.5 EXPRESSION OF *DKC1* ISOFORMS DURING *IN-VITRO* CELL DIFFERENTIATION

Since in previous semi-quantitative RT-PCR experiments the level of expression of the various *DKC1* transcripts appeared to be differentially modulated in the diverse tissues, I checked their expression during *in-vitro* differentiation of cultured cells. The expression levels of the various isoforms was then followed during differentiation of two cell lines by qRT-PCR. The expression of *DKC1* isoforms was firstly checked in Caco-2, a colon carcinoma cell line that differentiate spontaneously during cell culture, and subsequently in SK-N-BE(2), a neuroblastoma cell line whose differentiation is induced by *trans* retinoic acid. Caco-2 cells were grown as described (Pinto et al. 1983; Pignata et al. 1994) and cellular RNA extracted and analyzed by qRT-PCRs on days 4 (used as reference point), 6, 8, 11 and 15 of cell growth. In these experiments, the level of *S100G* mRNA, a gene encoding a calcium binding protein whose expression represents a differentiation marker of Caco-2 cells (Wang et al. 2004), was measured as positive control. As shown in Fig. 13, *S100G* expression increased significantly from day 6 onwards, peaking at day 11. In contrast, expression of *DKC1* isoforms 1, 4 and 7 slightly, but constantly decreased along differentiation, starting from day 6 onwards. Expression of isoforms 3, 5, 6 was reduced at a lesser extent, possibly suggesting a vaguely distinct conduct. However, it can reasonably be concluded that differentiation of Caco-2 cells is accompanied by only a modest reduction of all *DKC1* transcripts.



**Figure13. Expression of *DKC1* transcripts during Caco-2 cells differentiation.**  
Expression of the *S100G* gene was followed as differentiation marker. D: days.

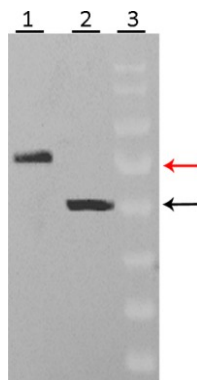
The expression of *DKC1* isoforms was then followed during differentiation of the human neuroblastoma SK-N-BE(2) cell line (Tahira et al. 1991; D'Alessio et al. 1995; Corvi et al. 1997). Total RNA was extracted at 0, 2, 4 and 6 days after *trans* retinoic acid induction, and the level of *RET* mRNA followed as positive control of cell differentiation (Tahira et al. 1991). As shown in Fig. 14, *RET* expression was strongly induced from day 2 onwards. The accumulation of *DKC1* transcripts appear instead to be differently regulated along cell differentiation. In fact, Isoform 1 expression seemed essentially unaffected, while Isoform 3 and 7 were progressively down-regulated. In contrast, Isoforms 6 and 4 were up-regulated, with Isoform 4 particularly and significantly induced. Finally, Isoform 5 showed a very weak expression in this cell line, so that its expression was not significantly detected at all time point. These results clearly indicated that expression of *DKC1* isoforms can be complex and finely tuned, and can be autonomously and differentially regulated during *in-vitro* cell differentiation. This finding suggests the different isoforms may exhibit different temporal and tissue expression profile also *in-vivo*. Thus, it would be interesting to analyze a wider panel of differentiating cells, in order to get a more complete picture of the expression profile of each isoform and acquire more information on their specific biological roles.



**Figure 14. Expression of *DKC1* transcripts during SK-N-BE(2) cell differentiation.**  
The expression of the *RET* gene was followed as differentiation marker .N.I: Not induced; D: days.

#### 4.6 OVEREXPRESSION OF ISOFORM 1 AND 7 IN TRANSIENTLY TRANSFECTED HeLa CELLS

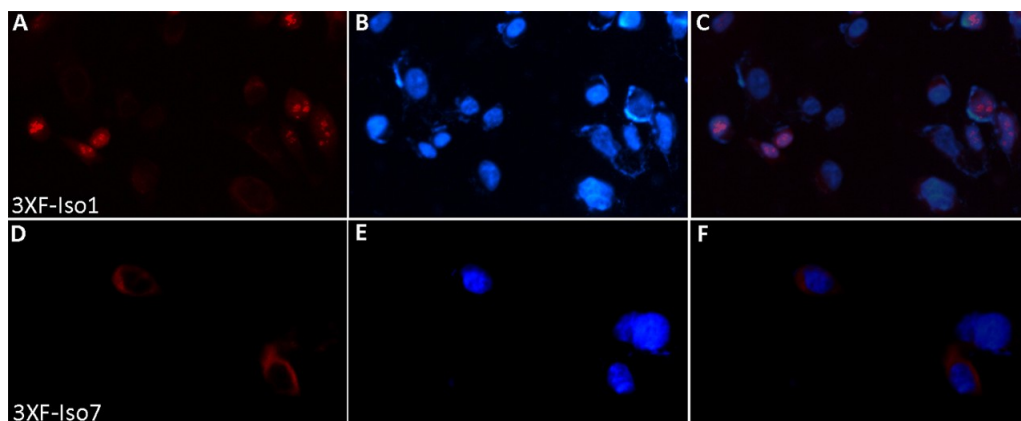
As mentioned above, my attention was mainly attracted by Isoform 7, on the basis of its coding properties and of the possibility that it may represent the main transcript expressed in a male child patient described in a recent case report (Pearson et al. 2008). To gain further information on its biological roles, I attempted to overexpress this isoform in HeLa cells. In parallel, I attempted to overexpress the canonical dyskerin (Isoform 1) as side-control. The CDS of these two isoform was then amplified by the use of specific primers and then cloned *in frame* into the mammalian expression vector p3XFlag-CMV-10. After suitable controls, each plasmid was used in transfection experiments, together with the p3XFlag-CMV-10 empty vector used as mock-control (here referred as 3XF-mock). Thereafter, I will refer to the p3XFlag-CMV-10 plasmid expressing Isoform 1 as 3XF-Iso1, and to the p3XFlag-CMV-10 plasmid expressing Isoform 7 as 3XF-Iso7. After transfection, clones transiently expressing 3XF-Iso1, 3XF-Iso7 and 3XF-mock were isolated and expression of isoforms 1 and 7 checked by western blot analysis. Recombinant proteins with sizes compatible with that expected (60 kDa for Isoform 1 and 50 kDa for Isoform 7) were efficiently revealed by the anti-Flag antibody (Fig. 15) in western analysis. The specific bands showed only a slight increase in the apparent weight, as usually observed also for the endogenous dyskerin. Once verified the effective expression of the *DKC1* isoforms, the transfected cells were grown on cover slips to visualize the intracellular protein localization by fluorescent immunolocalization.



**Figure 15. Western blot analysis of transiently transfected cells overexpressing isoforms 1 and 7.**

The anti-Flag antibody was used for immunodetection of the fused proteins. Lane 1: 3XF-Iso1, Lane 2: 3XF-Iso7, Lane 3: Page ruler protein marker (Fermentas). The red arrow indicates position of the 70 kDa band of the marker, the black arrow that of the 55 kDa band.

As expected, cells overexpressing the Isoform 1-FLAG fused protein showed the fluorescent signal specifically localized into the nucleoli (Fig. 16). This observation testified that addition of the FLAG peptide does not cause *per se* any mislocalization of the fused protein. In contrast, localization of the Isoform 7-FLAG protein was found totally cytoplasmic, with no nuclear signal detected (Fig. 16). This finding indicates that presence of the amino-terminal NLS is not sufficient to determine a nuclear localization of the protein, in good agreement with data reported by Heiss et al (1999). These authors prepared several deletion constructs encoding truncated forms of dyskerin and then examined their subcellular localization. Consistent with our data, a truncated protein lacking the C-terminal NLS (named dys446 $\Delta$ 514), even if 25 aa longer than Isoform 7, similarly showed a markedly reduced nuclear accumulation.



**Figure 16. Fluorescent immunolocalization of transiently transfected HeLa cells.**

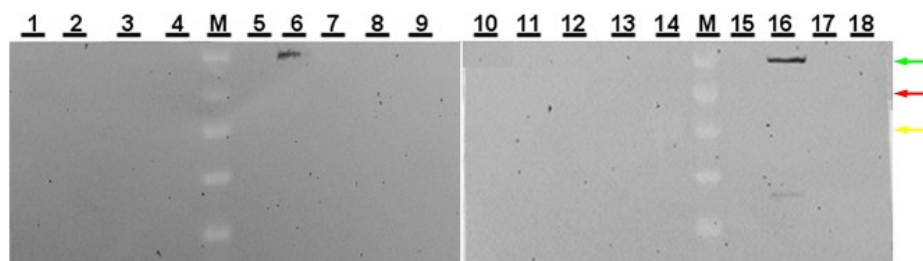
In A, immunolocalization of Isoform 1-FLAG fused protein, obtained with the anti-FLAG antibody. The signal localizes specifically in the nucleoli. In B, nuclei are counterstained with DAPI. C, merge of A and B. D: immunolocalization of Isoform 7-FLAG fused protein, obtained with the anti-FLAG antibody. The signal is totally absent from the nuclei and totally restricted to cytoplasm. In E, nuclei are counterstained with DAPI. F: merge of D and E.

Given that dyskerin is an essential component of two important nuclear complexes, such as that of the H/ACA snoRNPs and active telomerase, the potential role played by this cytoplasmic variant protein is quite puzzling. Nonetheless, previous reports raised the possibility that dyskerin might play some unexplained roles also in the cytoplasm. For example, in rat cells the dyskerin homolog protein, named Nap57p, has been reported to be involved in nucleo-cytoplasmic shuttling (Meier and Blobel 1994). More recently, proteomics strategies identified human dyskerin as a RPS19 interacting protein in cytoplasmic extracts (Orrù et al. 2007). Finally, a role for dyskerin in

directing cap-independent translation starting at the IRES sequenced has been assessed, but so far no clearly explained, in both mouse and human cells (Yoon et al. 2006, Bellodi et al. 2010, Montanaro et al. 2010) Although further experiments are required to establish whether Isoform7 may be directly involved in some of the above mentioned processes, it is relevant to note that this variant protein still keeps the PUA domain, which is essential to bind and stabilize the H/ACA snoRNAs, where the majority of the mutations causing dyskeratosis congenita cluster. Presence of the PUA domain is compatible with the hypothesis that this protein may have a role in cap-independent translation, or act as snoRNA chaperone, possibly transporting the snoRNA-derived miRNAs from the nucleus to the cytoplasm. According to this latter hypothesis, mammalian pseudouridine synthases might, as recently described in *Drosophila*, be involved in crucial developmental processes, such as *Notch* signaling, cell adhesion and cell competition (Tortoriello et al. 2010), thus interlacing cell growth and development.

#### 4.7 ISOLATION AND CHARACTERIZATION OF STABLY TRANSFORMED CLONES OVER-EXPRESSING DKC1 ISOFORM 7

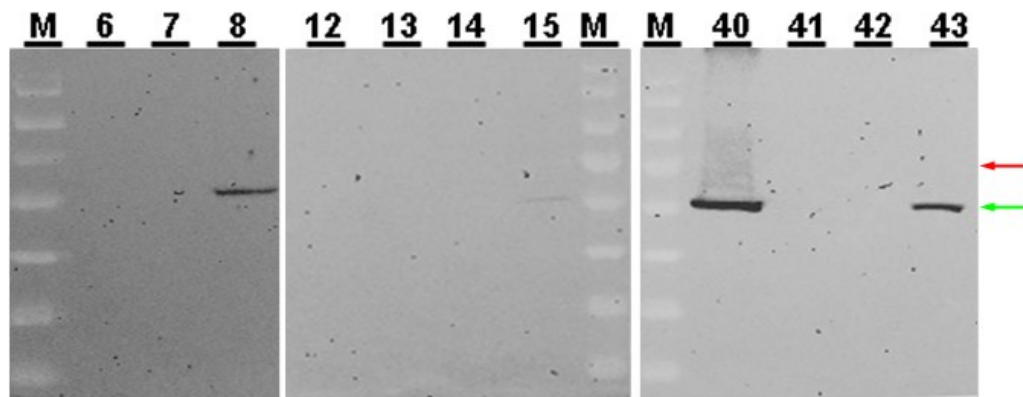
In order to carefully evaluate the effects caused by Isoform 7 overexpression, I then isolated and characterized stably transformed HeLa clones. In parallel, I also attempted to isolate stably transformed clones overexpressing Isoform 1. However, none of the selected 3XF-Iso1 clones showed expression of the fused protein. Only in two independently isolated clones (number 8 and 16, see Fig. 17), the anti-FLAG antibody detected a band of the apparent weight of about 100 kDa. This band is heavier than expected and possibly derives from an internal rearrangement of the expression construct which eliminated, at least in part, Isoform 1 sequences. In fact, I could not be able to amplify from the genomic DNA of these two clones Isoform 1-FLAG fused sequences. The failure to isolate clones overexpressing dyskerin has been already reported by other authors, and it has been related to so far unexplained toxic effects possibly exerted by abnormal accumulation of this protein (Ashbridge et al. 2009). In fact, successful overexpression of a member of this protein family has been to date obtained only for archaeal Cbf5p protein from *Pyrococcus furiosus* (Rashid et al. 2006).



**Figure 17. Western blot analysis of 3XF-Iso1 selected clones.**

The anti-Flag antibody was used for immunodetection. Numbers on the top mark different independent clones. Lanes M: Page ruler ladder (Fermentas). Colored arrows mark the position of 100 kDa (green), 70 kDa (red) and 40 kDa (yellow) bands of the marker. None of the examined clones expressed a fused protein product of the correct size.

The selection of 3XF-Iso7 stably transformed clones was instead successful, although only 10% of screened clones turned out to be positive. In total, I obtained four independent clones to submit to further analyses (see Fig. 18).



**Figure 18. Western blot analysis of 3XF-Iso7 selected clones.**

The anti-Flag antibody was used for immunodetection. Numbers on the top mark independently isolated clones. Lanes M: Page ruler ladder. Colored arrows mark the position of 70 kDa (red) and 40 kDa (yellow) bands of the marker. Clones 8,15, 40 and 43 all show expression of a protein exhibiting the correct size of the iso7-FLAG fused protein.

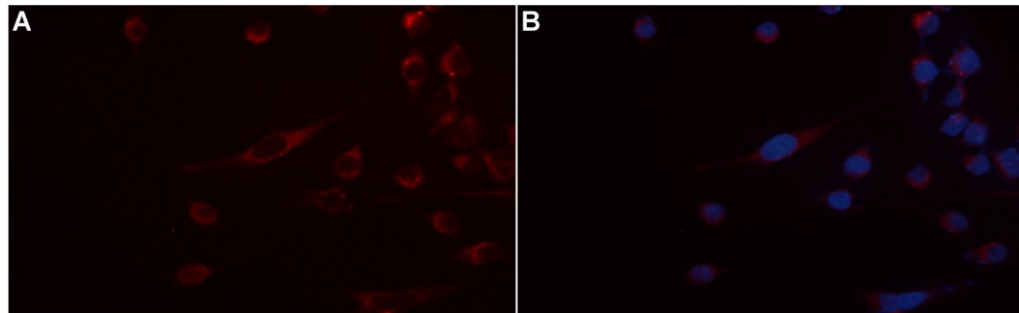
As expected, none of the 3XF-mock clones was positive by western analysis. Different clones were pooled to get a mean positional effect and used as control in the subsequent experiments.

#### **4.8 MORPHOLOGIC AND CELL-MATRIX ADHESION PROPERTIES OF CELLS OVEREXPRESSING ISOFORM 7**

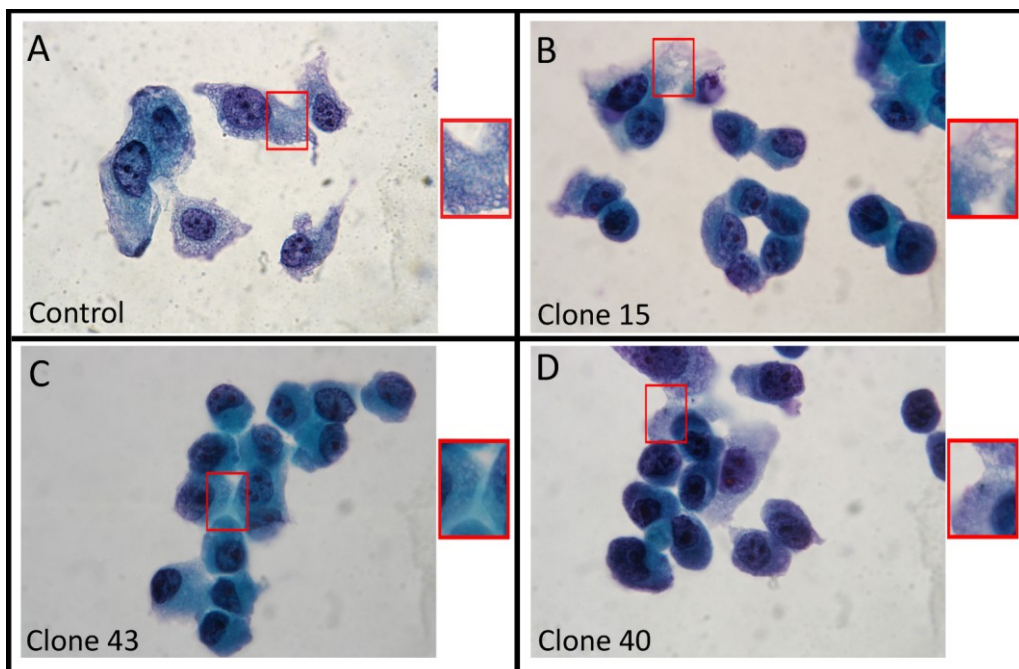
In order to address the biological function -if any- played by Isoform 7, stably transformed cells were analyzed for growth, morphology and adhesion properties. Four independently isolated clones over-expressing Isoform 7 at varying levels (see Fig. 18) were analyzed in these experiments. As expected, all these clones showed cytoplasmic localization of the fused protein (see Fig. 19 for an example), as previously observed in transiently transfected cells. Considering that dyskeratosis is a disorder accompanied by abnormal or premature epithelial keratinization, a preliminary inspection of the overall cell morphology was performed coloring cells attached to glass by Papanicolaou staining, which enables differential staining of nuclei (purple) to normal (blue/green) or keratinized cytoplasm (orange). However, no significant keratin increment was detected. Instead, I noticed that the transformed cells were invariantly characterized by a significant reduction in the number and the size of vacuole-like structures respect to the HeLa control cells (Fig. 20) which cytoplasm appears to be highly populated by these structures. Moreover, in each transformed clone the cells exhibit an increased tendency to aggregate to each other (see Figures 20, 21, 22). When the transformed cells were visualized under phase contrast microscopy, they again showed a tendency of adhere to each other when compared to controls, but also exhibited a different morphology, appearing more rounded and less elongated (Fig. 21). Intriguingly, changes in morphology and cell adhesion were



accompanied by an increased growth rate, indicating that the transformed cells also exhibit a growth advantage (see Fig. 21).



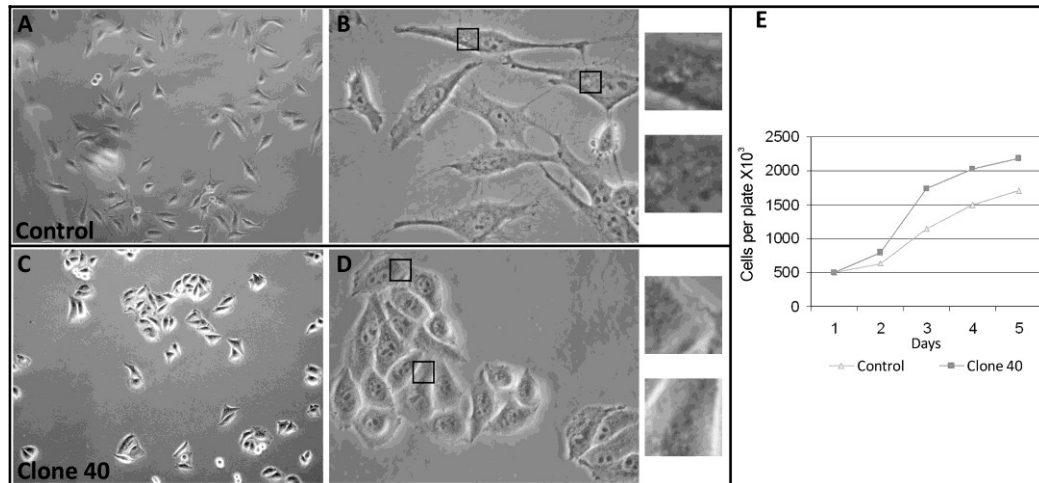
**Figure 19. Immunolocalization of Isoform 7 in cells from the stably transfected Clone 40.** In A, immunolocalization of Isoform 7-FLAG fused protein. The signal is absent from the nuclei and restricted to cytoplasm. In B, merge of A and nuclear counterstaining with DAPI.



**Figure 20. Papanicolaou staining of 3XF-Iso7 transformed and 3XF-mock control cells.** Typical fields from three 3XF-Iso7 independently transformed clones (15, 43 and 40) are shown. In each clone, reduced cytoplasm vacuolization was observed (see also red square areas, enlarged in right insets). Images are at 1000X magnification.

In order to investigate more deeply features typical of these transformed cells, sections of cell blocks were stained with Hematoxylin and Eosin, one of the most common method to detect a broad range of cytoplasmic, nuclear, and extracellular matrix features. Hematoxylin has a deep blue-purple color and stains nucleic acids; eosin is pink and stains proteins nonspecifically. In a typical tissue, nuclei are stained blue, whereas the cytoplasm and extracellular matrix have varying degrees of pink staining. This stain discloses abundant structural information, with specific functional implications. When 3XF-Iso7 cells from independent clones were stained, I observed presence of stretches of

adhesed cells forming localized tracts of structured pseudo epithelia that were resistant to Versene and trypsin treatments (Fig. 22). The formation of these pseudo epithelial rows, never detected in 3XF-mock cell, was again suggestive of increased cell adhesivity. Worth noting, this feature is well compatible with recent observations that silencing of members of the *DKC1* gene family is accompanied by reduced cell-cell adhesion (Sieron et al. 2009; Tortoriello et al. 2010).



**Figure 21. Phase contrast pictures of stably transfected cells.**

In A and B control cells; in B, cell areas rich in vacuole-like structures are black boxed. In 3XF-Iso7 cells (C and D) vacuole-like structures are strongly reduced; cells from Clone 40 are shown as an example. Note the tendency of transformed cells to grow in clusters. Pictures are at 400X (left) and 1000X (right) magnification. In E, a comparison of growth curves of 3XF-mock and 3XF-Iso7 (Clone 40) cells.

Careful examination of 3XF-Iso7 stained cells also revealed that they were characterized by a reduction of the nuclear-cytoplasmic ratio (N/C). Nuclear and cytoplasmic areas from 3XF-Iso7 transformed and 3XF-mock control cells were measured on acquired images by the ImageJ software. Statistical significance of this reduction was confirmed by submitting the values of nuclear and cytoplasmic areas from transformed and control cells to one-tailed Student's t-test, choosing  $p < 0.01$  as threshold. As shown in Table 4, the N/C ratio drops off in each independently transformed clone -although at various degrees- as consequence of a cytoplasmic enlargement, which is accompanied by invariance of the nuclear size. Cell differentiation is often accompanied by changes in N/C, with the nucleus retaining its dimension and the cytoplasm increasing in size. Conversely, actively proliferating cells show an increase of N/C ratio. The higher growth speed displayed by 3XF-Iso7 transformed cells would thus contrast with the N/C drop observed. However, cytoplasmic swelling has been correlated to higher growth speed also in *ras*-activated cells, where an enhanced proliferative activity has been reported to be accompanied by a significant increase in the cytoplasmic volume (Dartsch et al. 1994). Although this aspect needs further examination, it is plausible that a set-point shift of cell volume regulation, with subsequent cytoplasm swelling, might be

due to a rearrangement of the cytoskelatal architecture, of calcium metabolism or of volume-regulatory ion transporters, as suggested for *ras*-activated cells.

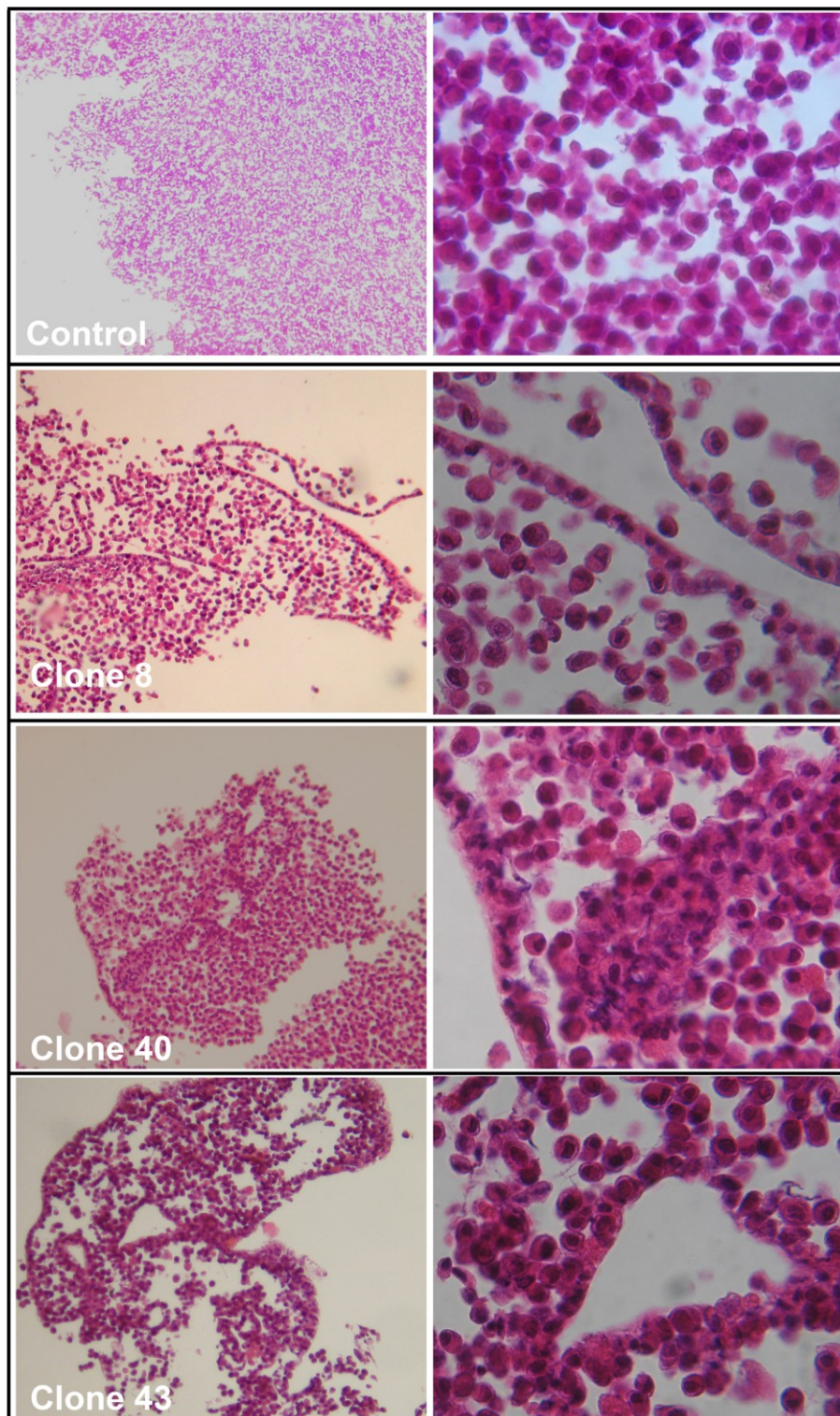
**Table 4. Nuclear-cytoplasmic ratio (N/C) of 3XF-Iso7 transformed and 3XF-mock control cells**

Cell	Nuclear area ( $\mu\text{m}^2$ )	$\pm\text{SD}$ ( $\mu\text{m}^2$ )	Cytoplasmic area ( $\mu\text{m}^2$ )	$\pm\text{SD}$ ( $\mu\text{m}^2$ )	N/C	$\pm\text{SD}$
<b>3XF-mock</b>	22.97	7.521	45.19	13.98	0.5515	0.2343
<b>Clone 8</b>	24.73	9.218	61.26*	23.16	0.4420*	0.2026
<b>Clone 15</b>	24.72	8.034	58.87*	16.31	0.4414*	0.1701
<b>Clone 40</b>	23.58	6.926	60.51*	19.14	0.4223*	0.1734
<b>Clone 43</b>	24.23	7.515	56.86*	14.95	0.4589*	0.2114

At least 100 cells of each sample were analyzed. \* Data showing a  $p < 0.01$  respect to control by one-tailed Student's t-test.

Worth noting, AgNOR staining, commonly used to highlight active Nucleolar Organizing Regions, did not evidentiate any difference in nuclear size and shape, or in the size or the number of the nucleoli (Fig. 23), further supporting the general view that the main alterations of transformed cells were restricted to the cytoplasm.

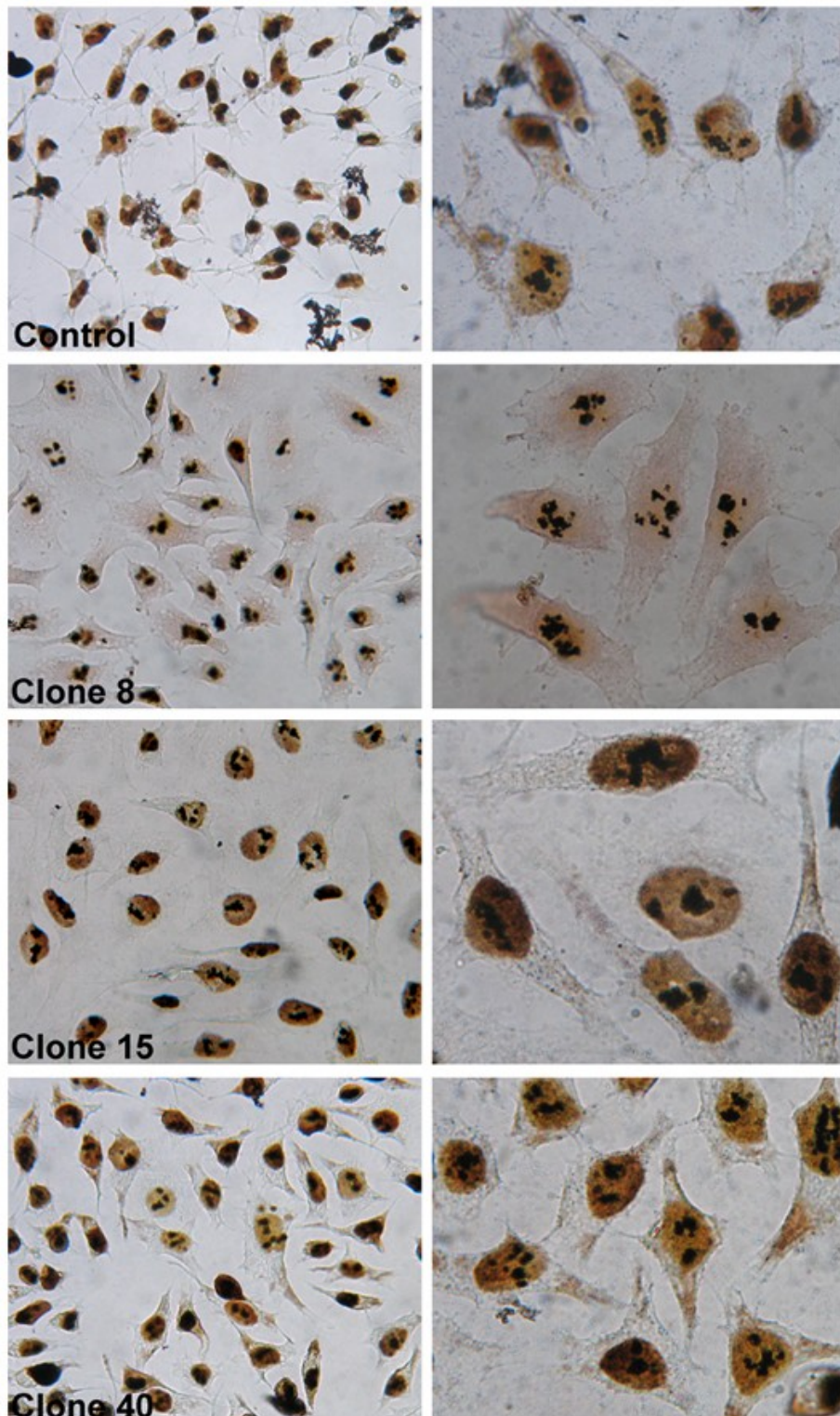
Finally, to further check cell adhesion properties, I performed a cell matrix adhesion assay followed by Crystal Violet staining. In this test, a significantly greater number of transformed cells quickly adhered to the culture substratum respect to the control (Fig. 24), and an increased tendency of cell-cell aggregation was observed. All the data obtained are thus consistent with the conclusion that Isoform 7 expression influences both cell-cell and cell-substratum adhesion. Intriguingly, a role for *DKC1* in cell adhesion has been recently noticed by Sieron et al. (2009), which reported that after RNAi-mediated gene silencing the cells were characterized by inability to reattach after trypsinization, or even spontaneous detachment from the substratum. In keeping with these observation, our laboratory recently reported that *in-vivo* RNAi silencing of the *DKC1* Drosophila homolog, the *mfl/Nop60b* gene, similarly causes lack of adhesion and defective cell-cell interactions (Tortoriello et al. 2010). The data presented here not only confirm a role for *DKC1* in cell adhesion, but also unveiled that the novel Isoform 7 is likely to participate to this function. Considering the peculiar cytoplasmic localization of this variant protein, it is plausible to postulate that it may mediate these effects by participating to the IRES-dependent translation or, alternatively, to the cytoplasmic transport of snoRNA-derived miRNAs.



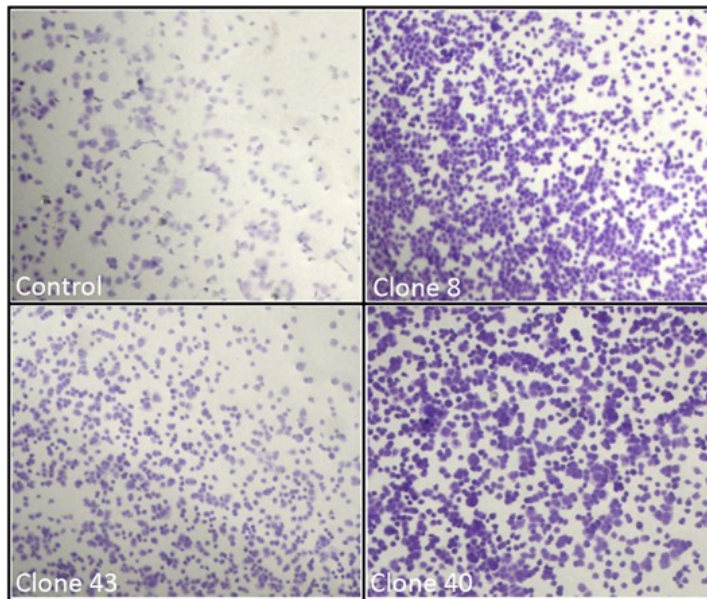
**Figure 22. Hematoxylin Eosin staining of 3XF-Iso7 transformed and 3XF-mock control cells.**

Typical fields from three 3XF-Iso7 independently transformed clones (8, 40 and 43) are shown. In each clone, structured rows of pseudo epithelia resistant to Versene treatment were observed. Images are at magnification 100X (left) and 400X (right).





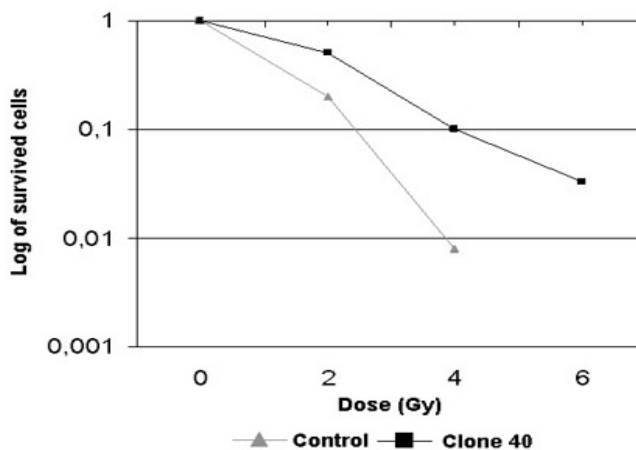
**Figure 23. AgNOR staining of 3XF-Iso7 transformed and 3XF-mock control cells.**  
 Typical fields from three 3XF-Iso7 independently transformed clones (8, 15 and 40) are shown. No significant variation in the nuclear (brown/yellow) size or in the nucleolar number or activity was observed in the transformed cells. Images are at magnification 400X (left) and 1000X (right).



**Figure 24. Cell matrix adhesion assay of 3XF-Iso7 transformed and 3XF-mock control cells.** Typical fields from three 3XF-Iso7 independently transformed clones (8, 43, and 40) are shown. In each clone, a significantly greater number of transformed cells adhered to the substratum respect to the control. Increased cell-cell aggregation was also observed. Images are at magnification 100X.

#### 4.9 *SURVIVAL OF 3XF-Iso7 TRANSFORMED CELLS AFTER X-ray TREATMENT*

X-ray treatment is known to generate DNA damages, which in part are repairable and quickly disappear, and in part are non-repairable, and thus persist in the cell population with detrimental effects. The non-repairable component depends on the dose of X-rays and on its subministration rate, since a higher rate corresponds to a higher damage. Finally, damage accumulation also depends on the proliferative state of the cells, since fast proliferating tissues are more sensitive than slower proliferating ones. Given that dyskerin loss-of-function mutations have recently been correlated to abnormal response to X-ray like damages (Gu et al. 2008), I decided to check the X-ray response of 3XF-Iso7 transformed cells. Unexpectedly, the dose-response curve to X-ray treatment indicated that these cells exhibit an increased repairing potential respect to controls, despite the fact that their growth rate was found significantly higher (Fig. 25).

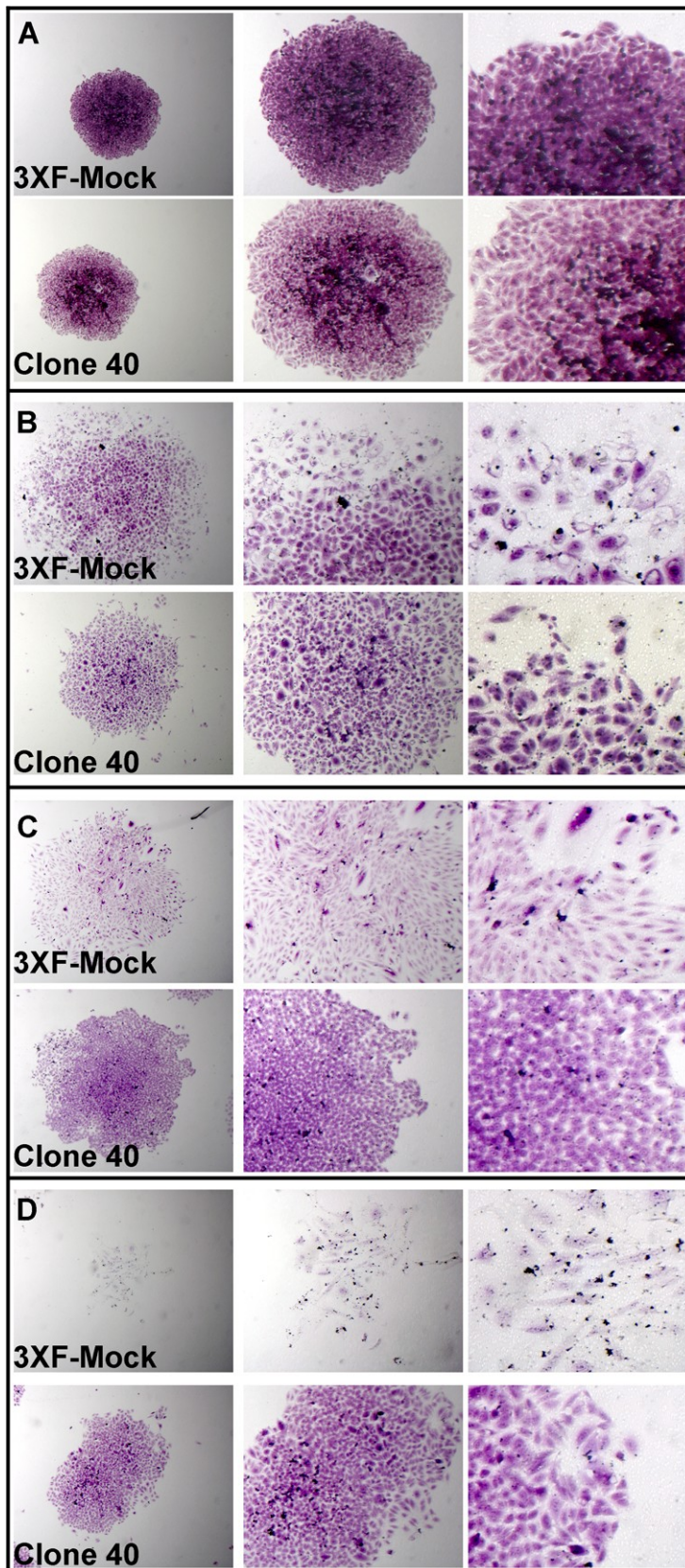


**Figure 25. X-ray dose response curve of 3XF-Iso7 transformed and 3XF-mock control cells.**

In the curve, the logarithm of the number of survived cells is plotted against the administered dose

Transformed and control cells were then analyzed by Crystal Violet staining after various doses of X-ray irradiation, i.e. after exposure to 0, 2, 4 and 6 Gy. At the 0 dose, a similar colony forming potential was observed for transformed and control cells, and the clones formed were of similar size. After the 2 Gy treatment, the transformed cells showed a higher percentage of survival and were mostly healthy, whereas control cells were often vacuolated and apoptotic and mixed to “ghosts” remnants of dead cells. A 4 Gy treatment further reduced the density of control clones, whereas the transformed clones were still dense but formed by smaller cells, possibly because a high proliferation rate. After a 6 Gy treatment, the density of transformed clones was still essentially similar to that of non treated cells. In contrast, control clones hardly survived and, when present, were very small, being composed of a few and poorly stained cells. These data revealed that Isoform 7 exerts a strong effect on X-ray response, and possibly on the response to other genotoxic stresses. This result is consistent with the previous finding of a *DKC1* involvement in the response to X-ray-like damages (Gu et al. 2008). The specific cytoplasmic localization exhibited by isoform 7 makes its role in the DNA damage response even more puzzling, and hints at a less direct role in the DNA repairing mechanisms.





**Figure 26. Crystal violet staining of 3XF-Iso7 transformed and 3XF-mock control clones after X-ray treatment.**

A) 0 Gy reference point.  
 B) Clone survival after 2 Gy treatment.  
 C) Clone survival after 4 Gy treatment.  
 D) Clone survival after 6 Gy treatment. Images are at magnification 25X (left), 50X (middle) and 100X, (right).



## 5 CONCLUSIONS

The results reported in this thesis revealed that transcription of the human *DKC1* gene is much more complex than generally accepted. Starting from *DKC1*-related ESTs annotated in the GenBank database, I defined the structure of five novel alternatively spliced transcripts and validated their effective expression in various cell lines and normal tissues. From the structural point of view, all the novel splice variants were characterized by retention of specific introns and presence of PTCs. However, all escape the NMD mechanism of RNA surveillance, supporting the view that they may play specific biological roles. This hypothesis is further supported by the finding that they can exhibit different expression profiles during the *in-vitro* differentiation of SK-N-BE(2) cells. Interesting to note, the production of isoforms 6 and 7 may antagonize, respectively, the biogenesis of the *DKC1* intron-encoded snoRNAs ACA36 and ACA56. This may have functional relevance, since both snoRNAs guide pseudouridylation of specific rRNA sites, and ACA56 can also act as miRNA precursor. Strikingly, the novel splice variants, with the exception of Isoform 7, all contain short upstream CDSs that would miss the C-terminal dyskerin domains, followed by significantly long *in frame* CDSs which, conversely, would miss the N-terminal functional domains. This general picture is consistent with the hypothesis that production of *DKC1* alternative isoforms may finely regulate the multiple functional roles played by this essential gene.

Among the novel alternative transcripts, I selected the Isoform 7 for further analyses. This isoform attracted my attention mainly on the basis of its coding properties and the possibility that, according to a recent case report, it should represent the main *DKC1* transcript expressed in a male child patient. To gain further information on its biological role, this isoform was overexpressed in stably transformed HeLa cells. Unexpectedly, the analysis of four independently isolated clones revealed that it encodes a variant protein with a strictly cytoplasmic localization. Given that dyskerin is an essential component of two important nuclear complexes, such as the H/ACA snoRNPs and the active telomerase, the role played by this cytoplasmic variant is quite puzzling. However, it is plausible that it might participate to ribosome assembly, mRNA translation or, alternatively, to cytoplasmic chaperoning of snoRNA-derived miRNAs.

While these aspects remain to be defined, characterization of Isoform 7 overexpressing clones demonstrated that they exhibit a higher growth rate, increased cell-cell and cell-substratum adhesion, and a marked higher tolerance to X-ray induced genotoxic damages respect to controls. These results are in agreement with functional data deriving from animal model system, and indicate that a more exhaustive functional analysis of the *DKC1* gene is necessary to unveiling how it may affect a so wide repertoire of fundamental cell processes.

## 6 ACKNOWLEDGMENTS

## 7 REFERENCES

- Alawi F and Lin P. Loss of dyskerin reduces the accumulation of a subset of H/ACA snoRNA-derived miRNA. *Cell Cycle* 2010;9:2467-69
- Amrani N, Sachs MS, Jacobson A. Early nonsense: mRNA decay solves a translational problem. *Nature Reviews Molecular Cell Biology* 2006;7:415-25
- Armanios M, Chen J-L, Chang Y-PC, Brodsky RA, Hawkins A, Griffin CA, Eshleman JR, Cohen AR, Chakravarti A, Hamosh A, Greider CW. Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proceedings of the National Academy of Sciences of the United States of America* 2005;104:15960-64
- Arnez JG and Steitz TA. Crystal Structure of Unmodified tRNA<sup>Gln</sup> Complexed with Glutaminyl-tRNA Synthetase and ATP Suggests a Possible Role for Pseudo-Uridines in Stabilization of RNA Structure. *Biochemistry* 1994;33:7560-67
- Ashbridge B, Orte A, Yeoman JA, Kirwan M, Vulliamy T, Dokal I, Klennerman D, Balasubramanian S. Single-Molecule Analysis of the Human Telomerase RNA-Dyskerin Interaction and the Effect of Dyskeratosis Congenita Mutations. *Biochemistry* 2009;48:10858-65
- Bancroft J, Gamble M, editors. *Theory and Practice of Histological Techniques*. Philadelphia:Churchill Livingstone; 2007.
- Barth S, Hury A, Liangù X-h, Michaeli S. Elucidating the Role of H/ACA-like RNAs in trans-Splicing and rRNA Processing via RNA Interference Silencing of the *Trypanosoma brucei* CBF5 Pseudouridine Synthase. *The Journal of Biological Chemistry* 2005;280:34558-68
- Bellodi C, Kopmar N, Ruggero D. Dereglulation of oncogene-induced senescence and p53 translational control in X-linked dyskeratosis congenita. *EMBO Journal* 2010;29:1865–76
- Bryan HG and Nixon RK. Dyskeratosis Congenita and Familial Pancytopenia. *JAMA* 1965;192:203-08
- Cavaillé J, Buiting K, Kieffmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie J-P, Brosius J, Hüttenhofer A. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proceedings of the National Academy of Sciences of the United States of America* 2000;26:14311-16
- Cohen SB, Graham ME, Lovrecz GO, Bache N, Robinson PJ, Reddel RR. Protein composition of catalytically active human telomerase from immortal

cells. *Science* 2007;315:1850-53

Corvi R, Savelyeva L, Schwab M. Patterns of oncogene activation in human neuroblastoma cells. *Journal of Neuro-Oncology* 1997;31:25-31

Costello MJ and Buncke CM. Dyskeratosis congenita. *Archives of Dermatology* 1956;73:123-32

Cuccurese M, Russo G, Russo A, Pietropaolo C. Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Research* 2005;33:5965-77

D'Alessio A, De Vita G, Cali G, Nitsch L, Fusco A, Vecchio G, Santelli G, Santoro M, de Franciscis V. Expression of the RET oncogene induces differentiation of SK-N-BE neuroblastoma cells. *Cell Growth and Differentiation* 1995;6:1387-94

Dartsch PC, Ritter M, Häussinger D, Lang F. Cytoskeletal reorganization in NIH 3T3 fibroblasts expressing the ras oncogene. *European Journal of Cell Biology* 1994;63:316-25

Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. A Human snoRNA with MicroRNA-Like Functions. *Molecular Cell* 2008;32:519-28

Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA Translation and Stability by microRNAs. *Annual Review of Biochemistry* 2010;79:351-79

Fogarty PF, Yamaguchi H, Wiestner A, Baerlocher GM, Sloand E, Zeng WS, Read EJ, Lansdorp PM, Young NS. Late presentation of dyskeratosis congenita as apparently acquired aplastic anaemia due to mutations in telomerase RNA. *Lancet* 2003;362:1628-30

Giordano E, Peluso I, Senger S, Furia M. minifly, a Drosophila Gene Required for Ribosome Biogenesis. *J Cell Biol.* 1999;144:1123-33

Gu B-W, Besslerù M, Mason PJ. A pathogenic dyskerin mutation impairs proliferation and activates a DNA damage response independent of telomere length in mice. *Proceedings of the National Academy of Sciences of the United States of America* 2008;105:10173-78

Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering* 1990;4:155-61

He J, Navarrete S, Jasinski M, Vulliamy T, Dokal I, Bessler M, Mason PJ.

- Targeted disruption of Dkc1, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice. *Oncogene* 2002;21:7740-44
- Heiss NS, Girod A, Salowsky R, Wiemann S, Pepperkok R, Poustka A. Dyskerin Localizes to the Nucleolus and Its Mislocalization Is Unlikely to Play a Role in the Pathogenesis of Dyskeratosis Congenita. *Human Molecular Genetics* 1999;8:2515-24
- Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genetics* 1998;19:32-38
- Jiang W, Middleton K, Yoon H-J, Fouquet C, Carbon J. An Essential Yeast Protein, CBF5p, Binds In Vitro to Centromeres and Microtubules. *Molecular and Cellular Biology* 1993;13:4884-93
- Jung C-H, Hansen MA, Makunin IV, Korbie DJ, Mattick1 JS. Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. *BMC genomics* 2010;11:77
- Kirwan M and Dokal I. Dyskeratosis congenita, stem cells and telomeres. *Biochimica Biophysica Acta* 2009;1792:371-79
- Kittur N, Darzacq X, Roy S, Singer RH, Meier UT. Dynamic association and localization of human H/ACA RNP proteins. *RNA* 2006;12:2057-62
- Knight SW, Vulliamy TJ, Morgan B, Devriendt K, Mason PJ, Dokal I. Identification of novel DKC1 mutations in patients with dyskeratosis congenita: implications for pathophysiology and diagnosis. *Human Genetics* 2001;108:299-303
- Kurnikova M, Shagina I, Khachatryan L, Schagina O, Maschan M, Shagin D. Identification of a Novel Mutation in DKC1 in Dyskeratosis Congenita. *Pediatr Blood Cancer* 2009;52:135-37
- Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100:189-92
- Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Reviews Molecular Cell Biology* 2004;5:89-99
- Marrone A, Walne A, Dokal I. Dyskeratosis congenita: telomerase, telomeres and anticipation. *Current Opinion in Genetics & Development* 2005;15:249-57

- Pinto M, Robine-Leon S, Appay MD, Keding M, Triadou N, Dussaulx E, Lacroix B, Simon-Assmann P, Haffen K, Fogh J, Zweibaum A. Enterocyte-like differentiation and polarization of the human colon carcinoma cell line Caco-2 in culture. *Biology of the Cell* 1983;47:323-29
- Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H. Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and Dyskeratosis Congenita. *Molecular Cell* 2006;21:249-60
- Richard P, Kiss AM, Darzacq X, Kiss T. Cotranscriptional Recognition of Human Intronic Box H/ACA snoRNAs Occurs in a Splicing-Independent Manner. *Molecular and Cellular Biology* 2006;26:2540-49
- Richard P and Kiss T. Integrating snoRNP assembly with mRNA biogenesis. *EMBO Report* 2006;7:590-92
- Ruggero D, Grisendi S, Piazza F, Rego E, Mari F, Rao PH, Cordon-Cardo C, Pandolfi PP. Dyskeratosis Congenita and Cancer in Mice Deficient in Ribosomal RNA Modification. *Science* 2003;299:259-62
- Sambrook J and Russel DW. *Molecular Cloning: a laboratory manual* New York: Cold Spring Harbor Laboratory Press/Elsevier; 2001. 2344p
- Saraiya AA and Wang CC. snoRNA, a Novel Precursor of microRNA in *Giardia lamblia*. *Plos Pathogens* 2008;4:e1000224
- Scoggins RB, Prescott KJ, Asher GH, Blaylock WK, Bright RW. Dyskeratosis congenita with Fanconi-type anemia: investigations of immunologic and other defects. *Clin. Res* 1971;19:409
- Sieron P, Hader C, Hatina J, Engers R, Wlazlinski A, Müller M, Schulz WA. DKC1 overexpression associated with prostate cancer progression. *British Journal of Cancer* 2009;101:1410-16
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanarajc TA, Soreq H. Function of alternative splicing. *Gene* 2005;344:1-20
- Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS. Small RNAs derived from snoRNAs. *RNA* 2009;15:1233-40
- Tahira T, Ishizaka Y, Itoh F, Nakayasu M, Sugimura T, Nagao M. Expression of the ret proto-oncogene in human neuroblastoma cell lines and its increase during neuronal differentiation induced by retinoic acid. *Oncogene* 1991;6:2333-38
- Tortoriello G, de Celis JF, Furia M. Linking pseudouridine synthases to

growth, development and cell competition. *The FEBS Journal* 2010;277:3249-63

Turano M., Angrisani A., De Rosa M., Izzo P., Furia M. Real-time PCR quantification of human DKC1 expression in colorectal cancer. *Acta Oncol.* 2008;47:1598-9

Vulliamy T, Beswick R, Kirwan M, Marrone A, Digweed M, Walne A, Dokal I. Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *Proceedings of the National Academy of Sciences of the United States of America* 2008;105:8073-78

Vulliamy TJ, Knight SW, Heiss NS, Smith OP, Poustka A, Dokal I, Mason PJ. Dyskeratosis congenita caused by a 3' deletion: germline and somatic mosaicism in a female carrier. *Blood* 1999;94:1254-60

Vulliamy TJ, Knight SW, Mason PJ, Dokal I. Very Short Telomeres in the Peripheral Blood of Patients with X-Linked and Autosomal Dyskeratosis Congenita. *Blood Cells, Molecules, and Diseases* 2001;27:353-57

Walne AJ, Marrone A, Dokal I. Dyskeratosis Congenita: A Disorder of Defective Telomere Maintenance? *Int J Hematol.* 2005;82:184-89

Walne AJ, Vulliamy T, Beswick R, Kirwan M, Dokal I. TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. *Blood* 2008;112:3594-600

Walne AJ, Vulliamy T, Marrone A, Beswick R, Kirwan M, Masunari Y, Al-Qurashi F-h, Aljurf M, Dokal I. Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Human Molecular Genetics* 2007;16:1619-29

Wang L, Klopot A, Freund J-N, Dowling LN, Krasinski SD, Fleet JC. Control of differentiation-induced calbindin-D9k gene expression in Caco-2 cells by cdx-2 and HNF-1 $\alpha$ . *Am J Physiol Gastrointest Liver Physiology* 2004;287:G943-G53

Yoon A, Peng G, Brandenburg Y, Zollo O, Xu W, Rego E, Ruggero D. Impaired Control of IRES-Mediated Translation in X-Linked Dyskeratosis Congenita. *Science* 2006;312:902-06

Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Molecular and Cellular Biology* 1999;19:7461-72

## **8 ORIGINAL PAPERS**

- nization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group. *J Clin Oncol* 1997; 15:1837-43.
- [7] Krarup-Hansen A, Helweg-Larsen S, Schmalbruch H, Rorth M, Krarup C. Neuronal involvement in cisplatin neuropathy: Prospective clinical and neurophysiological studies. *Brain* 2007;130:1076-88.
- [8] Vigliani MC, Magistrello M, Polo P, Mutani R, Chio A. Risk of cancer in patients with Guillain-Barré syndrome (GBS). A population-based study. *J Neurol* 2004;251:321-6.
- [9] Tho LM, O'Leary CP, Horrocks I, Al-Ani A, Reeds NS. Guillain-Barre syndrome occurring after adjuvant chemoradiotherapy for endometrial cancer. *Gynecol Oncol* 2006; 100:615-7.
- [10] Christodoulou C, Anastasopoulos D, Visvikis A, Mellan S, Detski I, Tsiakalos G, et al. Guillain-Barré syndrome in a patient with metastatic colon cancer receiving oxaliplatin-based chemotherapy. *Anti-Cancer Drugs* 2004;15:997-9.

## Real-time PCR quantification of human DKC1 expression in colorectal cancer

MIMMO TURANO<sup>1</sup>, ALBERTO ANGRISANI<sup>1</sup>, MARINA DE ROSA<sup>2</sup>, PAOLA IZZO<sup>2</sup> & MARIA FURIA<sup>1</sup>

<sup>1</sup>Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli "Federico II", via Cinthia, 80126 Naples, Italy,

<sup>2</sup>Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli Federico II, via S. Pansini 5, 80131 Naples, Italy

### To the Editor

Mutations in the highly conserved human DKC1 gene cause the rare genetic disease X-linked recessive dyskeratosis congenita (X-DC) [1]. X-DC patients display features of premature aging, mucosal leukoplakia, nail dystrophy, skin pigmentation, interstitial fibrosis of the lung, bone-marrow failure and increased susceptibility to cancer [2]. Hypomorphic DKC1 mutant mice recapitulate the major features of X-DC, including increased cancer susceptibility. These mice were in fact highly prone to tumors and developed a variety of them, most commonly from lung and mammary gland, indicating that DKC1 may act as an important tumor-suppressor *in vivo* [3]. DKC1 encodes a nucleolar protein, named dyskerin, which acts as pseudouridine synthase and constitutes one of the four core protein components of the specific H/ACA RNPs involved in RNA pseudouridylation [1]. Dyskerin proved to be essential also for proper rRNA processing, thus playing multiple roles on ribosome biogenesis [3]. Moreover, this protein has been recently identified as an essential component of the catalytically active human telomerase complex, together with the human telomerase reverse transcriptase (TERT) and the RNA component of telomerase (TERC) [4]. A quantitative analysis of dyskerin

mRNA expression was recently performed by real-time RT-PCR on a series of breast carcinomas [5]. In this study, the TERC levels and the overall degree of rRNA pseudouridylation were also evaluated. The amount of dyskerin mRNA was found to be variable, but always significantly associated with TERC and rRNA pseudouridylation levels.

Considering that dyskerin and TERT are both constituents of the active human telomerase enzyme complex [4] and that TERT has been indicated as a potential biomarker for colorectal cancer [6], we wished to check whether dyskerin and TERT mRNA levels would vary in parallel in colorectal tumors. According to the multiple role played by dyskerin in ribosome biogenesis, its expression levels were expected to be highly variable in different patients, possibly depending on age, sex and general metabolic conditions. To take in account this aspect, in our experiments we always referred as control to adjacent non-tumor mucosa matched samples.

Normal colorectal mucosa and colorectal cancer tissues were then sampled from 8 patients affected by sporadic colon cancer and assayed by quantitative real-time RT-PCR analysis. Institutional ethical approval and informed consent were given by the patients undergoing surgery for colorectal cancer. Total RNA was extracted purified and reverse-

Correspondence: Mimmo Turano, Dipartimento di Biologia Strutturale e Funzionale, Complesso Universitario di Monte S. Angelo, Università di Napoli Federico II, Via Cinthia 80126 Naples, Italy. Tel: +39 081 679076. Fax: +39 081 679300. E-mail: mimmo.turano@unina.it

(Received 11 December 2007; accepted 7 January 2008)

ISSN 0284-186X print/ISSN 1651-226X online © 2008 Informa UK Ltd. (Informa Healthcare, Taylor & Francis AS)  
DOI: 10.1080/02841860801898616

RIGHTS LINK  
Copyright Clearance Center



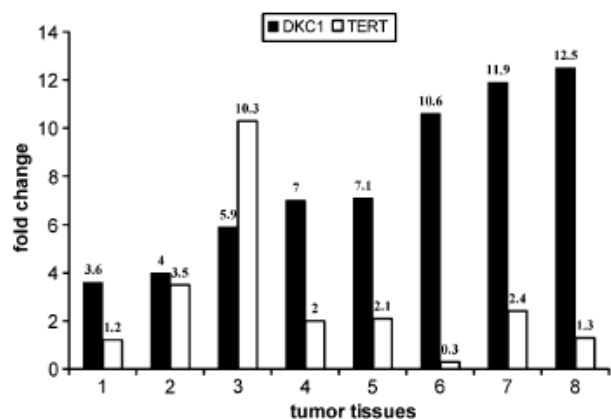


Figure 1. Expression level of TERT and DKC1 mRNA in colon tumor tissues compared to each matched normal mucosa. This experiment was performed three times with similar results. A typical experiment is shown.

transcribed; quantitative real-time PCR was then performed by using the iQ<sup>TM</sup> 5 Multicolor Real-Time PCR Detection System (Bio-Rad). All PCR reactions were carried out in a total volume of 15  $\mu$ l using 5 ng of cDNA, 5 pmol of each primer and 7.5  $\mu$ l of iQ<sup>TM</sup> SYBR Green Supermix 2X (Bio-Rad). All quantifications were normalized to the expression of large subunit of RNA pol II (POLR2A) endogenous control. PCR oligonucleotides, designed by using the Primer Express software (Applied Biosystems), were: h-TERT Ex-5/6-F tggacttggtcaaggtggatgtga; h-TERT Ex-6-R ttgatgatgtggtgcatga; h-DKC1 Ex-11-F ctgcgaagtgggttttaggt; h-DKC1 Ex-13-R gaggtgtgtgctgaagtgtgt; h-POLR2A Ex-22-F tgctccgtattcgcacatg; h-POLR2A Ex-23-R tccatctgtccaccacctctt. Quantitative PCR analysis was carried out using the  $2^{-\Delta\Delta C_T}$  method [7].

The results obtained, shown in Figure 1, indicated that TERT exhibited a variable mRNA level, although in most cases it was up-regulated respect to the matched controls ( $p = 0.066$ , one-sided  $t$  test). This observation essentially confirmed the variability of TERT mRNA levels recently described by Saleh et al. in colorectal tumor tissues respect to adjacent non-tumor mucosa [6]. In contrast to the fluctuation displayed by TERT levels, in our experiments we found that the amounts of dyskerin mRNA were always significantly higher in tumors than in normal tissues ( $p = 0.0004$ , one-sided  $t$  test), despite the fact

that the two proteins are equimolecular constituents of the active human telomerase enzyme complex [4]. This different behaviourReal-time PCR quantification of human DKC1 expression in colorectal cancer is likely due to the additional roles played by dyskerin as structural component of the H/ACA snoRNPs involved in rRNA processing, folding and modification. Indeed, increase of dyskerin mRNA levels is expected to be a general feature marking highly proliferating cells. In this light, the variable levels of DKC1 mRNA observed in breast carcinomas [5] might be explained by the fact that SW48 *in vitro* cultured cells, instead of adjacent non-tumor tissues, were utilised as calibrator. This way, the fluctuation of dyskerin mRNA level existing among different patients would have been ignored.

In conclusion, our data support the view that DKC1 expression may furnish a less variable and more sensible marker for proliferative colon disorders. Although analysis of a more exhaustive series of tumors is needed, the high concordance of the data obtained led us to suggest that evaluation of DKC1 mRNA level can provide a useful clue in the diagnosis and prognosis of colorectal cancer, and possibly of other proliferative disorders.

## References

- [1] Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* 1998;19:32–8.
- [2] Dokal I. Dyskeratosis congenita in all its forms. *Br J Haematol* 2000;110:768–79.
- [3] Ruggero D, Grisendi S, Piazza F, Rego E, Mari F, Rao PH, et al. Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* 2003;299:259–62.
- [4] Cohen SB, Graham ME, Lovrecz GO, Bache N, Robinson PJ, Reddel RR. Protein composition of catalytically active human telomerase from immortal cells. *Science* 2007;315:1850–3.
- [5] Montanaro L, Brigotti M, Clohessy J, Barbieri S, Ceccarelli C, Santini D, et al. Dyskerin expression influences the level of ribosomal RNA pseudo-uridylation and telomerase RNA component in human breast cancer. *J Pathol* 2006;210:10–8.
- [6] Saleh S, Lam AK, Ho YH. Real-time PCR quantification of human telomerase reverse transcriptase (hTERT) in colorectal cancer. *Pathology* 2008;40:25–30.
- [7] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods* 2001;25:402–8.





## A novel *Drosophila* antisense scaRNA with a predicted guide function

Giuseppe Tortoriello, Maria Carmela Accardo, Filippo Scialò, Alberto Angrisani, Mimmo Turano, Maria Furia \*

Department of Structural and Functional Biology, University of Naples "Federico II", Complesso Universitario Monte Santangelo, via Cinthia, 80126 Napoli, Italy

### ARTICLE INFO

#### Article history:

Received 12 November 2008  
Received in revised form 3 February 2009  
Accepted 3 February 2009  
Available online 20 February 2009

Received by C. Feschotte

#### Keywords:

*Drosophila*  
scaRNA  
snoRNA  
Cajal body  
snRNA U1b  
GAS5

### ABSTRACT

A significant portion of eukaryotic small ncRNA transcriptome is composed by small nucleolar RNAs. From archaeal to mammalian cells, these molecules act as guides in the site-specific pseudouridylation or methylation of target RNAs. We used a bioinformatics search program to detect *Drosophila* putative orthologues of U79, one out of ten snoRNAs produced by GAS5, a human ncRNA involved in apoptosis, susceptibility to cancer and autoimmune diseases. This search led to the definition of a list of U79-related fly snoRNAs whose genomic organization, evolution and expression strategy are discussed here. We report that an intriguing novel specimen, named *Dm46E3*, is transcribed as a longer, unspliced precursor from the reverse strand of *eiger*, a fly regulatory gene that plays a key role in cell differentiation, apoptosis and immune response. Expression of *Dm46E3* was found significantly up-regulated in a mutant strain in which *eiger* transcription is greatly reduced, suggesting that these two sense–antisense genes may be mutually regulated. Relevant to its function, *Dm46E3* concentrated specifically in the Cajal bodies, followed a dynamic spatial expression profile during embryogenesis and displayed a degenerate antisense element that enables it to target U1b, a developmentally regulated isoform of the U1 spliceosomal snRNA that is particularly abundant in embryos.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

In eukaryotic cells, many regulatory circuitries implicated in important aspects of development and cell differentiation involve the expression of small ncRNAs (reviewed by Mattick, 2004; Mattick and Makunin, 2006). Indeed, a significant portion of the eukaryotic small ncRNA transcriptome is composed by small nucleolar RNAs (snoRNAs). These molecules, generally ranging from 60 to 300 nucleotides in length, are present from Archaeal to mammalian cells. With a few exceptions, snoRNAs are classified into two major families, named box C/D and H/ACA, on the basis of common sequence motifs, structural features and common sets of associated proteins. A few members of each family are required for proper rRNA processing, whereas the majority directs, by base-pairing guiding mechanisms, the two most common types of nucleotide modifications present on eukaryotic RNAs (reviewed by Bachellerie et al., 2002; Kiss, 2002). H/ACA snoRNAs guide in fact RNA pseudouridylation, while C/D snoRNAs direct 2'-O-ribose methylation of target RNAs (Bachellerie et al., 2002; Kiss, 2002). Methylation guide snoRNAs display a simple structure, characterized by the presence of consensus C (5'-RUGAUGA-3') and D (5'-CUGA-3') motifs close to the 5' and 3' termini of the molecule,

respectively. Additional and often degenerated internal copies of C and D elements (designated C' and D') may also be present and may confer to the snoRNA the ability to guide methylation at two sites. In fact, D/D' upstream regions act as antisense elements able to select the residue to modify through the formation of specific duplexes with the RNA target (Bachellerie et al., 2002; Kiss, 2002). According to the so called D/D' +5 rule, the methylated site is invariably complementary to the snoRNA nucleotide mapping five positions upstream of the D/D' box. In the methylation process, the methyltransferase catalytic activity is furnished by fibrillarin, one of the four evolutionarily conserved core proteins that associate to snoRNAs of this family to compose the C/D functional small nucleolar ribonucleoprotein (snoRNPs) complexes (reviewed by Henras et al., 2004; Kiss et al., 2006). Initially, the role of snoRNAs was thought to be restricted to rRNA modification in ribosome biogenesis, but it is now evident that these molecules can target other RNAs, including snRNAs and mRNAs (Henras et al., 2004). Indeed, a subgroup of snoRNAs involved in snRNA modification do not localize in the nucleolus but in the Cajal bodies, nucleoplasmic organelles in which concentrate many ribonucleoprotein (RNP) factors whose major function is in maturation, assembly and/or trafficking of snRNPs, snoRNPs and transcription complexes (reviewed by Stanek et al., 2008; Pontes and Pikaard, 2008). Cajal body-specific RNAs (scaRNAs) can be of either the C/D or H/ACA type or can comprise a H/ACA domain embedded in a C/D box structure, and are typically involved in the methylation and/or pseudouridylation of the Pol II-transcribed snRNAs of the U1 spliceosome (Richard et al., 2003; Henras et al., 2004).

During the last few years, the repertoire of snoRNA's biological role was continuously widening. For instance, human C/D box snoRNAs

Abbreviations: ncRNA, non coding RNA; snoRNA, small nucleolar RNA; CB, Cajal body; scaRNA, small Cajal bodies RNA; CAB box, Cajal body-specific localization signal; snoRNP, small nucleolar ribonucleoprotein; dUHG, *Drosophila* homolog of human UHG; gas5, growth arrest-specific transcript 5; TMG, 2,2,7-trimethylguanosine cap; DIG, digoxigenin; JNK pathway, Jun N-terminal Kinase pathway.

\* Corresponding author. Tel.: +39 081679072 (office), +39 081/679071, +39 081 679076 (lab); fax: +39 081/679233.

E-mail address: [mfuria@unina.it](mailto:mfuria@unina.it) (M. Furia).



specifically expressed in the brain have been implicated in the interference with A-to-I editing (Vitali et al., 2005) and in the regulation of alternative mRNA splicing of the serotonin receptor (Kishore and Stamm, 2006). The human non-coding growth arrest-specific transcript 5 (GAS5), which encodes ten different snoRNAs in its introns (U44, U47, U74, U75, U76, U77, U78, U79, U80, U81; Smith and Steitz, 1998), was shown to play a critical role in the control of cell growth and apoptosis, to be down regulated in breast cancer (Mourtada-Maarabouni et al., 2009), and has evoked as a candidate gene in the development of autoimmune disease (Mourtada-Maarabouni et al., 2008). Together with the recent demonstration of tumor suppressor characteristics in the human snoRNA U50 (Dong et al., 2008), these observations suggest that snoRNA genes may be involved in controlling oncogenesis and sensitivity to therapy in cancer. In addition, an increasing number of orphan snoRNAs which lack antisense to rRNAs or snRNAs have been experimentally identified from different organisms. Since some of them display a developmentally or tissue specific pattern, and in mammals also imprinting, it has been suggested that they play still uncharacterized regulatory roles on gene expression (reviewed by Rogelj, 2006).

Genomic organization of C/D snoRNA genes has been found to be complex and variegated. In plant, yeast and protozoan the large majority of these genes is independently transcribed, whereas in nematodes monocistronic snoRNAs are almost as abundant as those intron-encoded (Huang et al., 2007). In *Drosophila*, organization and expression of C/D snoRNA genes have been systematically investigated. The two genome-wide computational approaches performed so far (Accardo et al., 2004; Huang et al., 2005) were essentially based on the primary structural elements of snoRNAs and the sequence complementarity to modified sites present on rRNA. However, these screens had to rely on the conservation of rRNA methylation sites described in other eukaryotes, since direct information on methylated residues present on the fruit fly rRNA is still missing. Nonetheless, they were effective and largely contributed to furnish an overall view of organization of snoRNA genes in the *Drosophila* genome. Similarly to vertebrates, most of fly C/D snoRNAs are intron-encoded and usually arranged in the mode of one snoRNA per intron (Yuan et al., 2003; Accardo et al., 2004; Huang et al., 2005), whereas only a very few are independently transcribed monocistrons. Altogether, bioinformatics and molecular approaches have so far led to annotation of 63 fly C/D box snoRNAs, which are predicted to account for about two-thirds of the rRNA methylations and a very few snRNAs modifications, suggesting that a substantial number of *Drosophila* snoRNA genes still await to be recognized.

In a previous analysis (Accardo et al., 2004), we used the SnoScan program (Lowe and Eddy, 1999) and sequences of yeast rRNA methylation sites conserved in *Drosophila* to search for methylation guide snoRNAs in the fruit fly genome. To extend further this approach, we performed a new screen using the same program but sequences of human rRNA methylation sites conserved in *Drosophila*. Among these conserved sites, we specifically selected for the analysis those that were methylated on human, but not on yeast rRNA. As a first outcome, we report here the results of our quest of U79-related fly snoRNAs. Human U79 is responsible for methylation of 28S at a site highly conserved in metazoans, but unmodified in yeast, and represents one out of ten snoRNAs encoded by the GAS5 pro-apoptotic ncRNA involved in cell growth, cancer and autoimmune diseases (Smith and Steitz, 1998; Nakamura et al., 2008; Mourtada-Maarabouni et al., 2008, 2009). This search led to the identification of two novel snoRNAs with peculiar genomic arrangement, one of which was antisense to *eiger* (*egr*), a *Drosophila* regulatory gene encoding a Tumor Necrosis Factor-like ligand involved in the regulation of cell differentiation, death and immunity (Igaki et al., 2002; Kaupila et al., 2003; Schneider et al., 2007). Besides adding further information on the repertoire, evolution and expression strategy of *Drosophila* C/D snoRNAs, the detailed characterisation of this antisense snoRNA

provides a glimpse into the molecular mechanisms underlying snoRNA biological functions, as well as those governing *egr* regulation.

## 2. Materials and methods

### 2.1. Fly strains

Canton S was used as wild-type strain in all experiments; the *egr*<sup>-3</sup> loss of function mutants was isolated by Igaki et al. (2002).

### 2.2. RNA analysis

Total RNA from embryos, larvae, adults and S2 cells was extracted using TRI Reagent® (Sigma) following manufacturer's instruction. RNA was treated with RQ1 DNase (Promega), phenol:chloroform extracted, precipitated with LiCl and resuspended with RNase free water. For Northern blot analysis, 6 µg of total RNA were electrophoresed and transferred onto Hybond-NX (Amersham) filters for hybridization. 0.3–0.5 kb genomic DNA fragments spanning the SnoScan detected sequence were used as specific probes; these probes were PCR-amplified on genomic DNA by using the appropriate primer pairs and <sup>32</sup>P-labelled using the Nick Translation Kit (Roche). The list of oligonucleotides utilised for Northern blotting or RT-PCR experiments is shown in Supplemental Table 1.

Basic cloning techniques, PCR amplification, DNA extraction, manipulation and labelling, screening and sequencing techniques were carried out according to Sambrook and Russell (2001). The size of RNAs was determined by using the High range and Low range RNA Molecular Weight Markers (Fermentas). In quantitative real-time RT-PCR analysis, the RNA (1 µg) was reverse transcribed using QuantiTect® Rev. Transcription (Qiagen) using manufacturer's condition and diluted 1:10. Quantitative real-time RT-PCR experiments were performed using iQ™5 Multicolor Real-Time PCR Detection System (Biorad). All PCR reactions were carried out in a final volume of 15 µl using 1 µl of diluted cDNA, 7.5 µl of 2× SYBR® GreenER™ (Invitrogen) and 5 pmol of each primer. Sequences of all utilised primers were designed using Primer 3 software (<http://frodo.wi.mit.edu/>) and are shown in Supplemental Table 1. 5.8 S rRNA was used as endogenous control for samples normalization. Quantitative PCR analyses were performed using the 2<sup>-ΔΔCT</sup> method (Livak and Schmittgen, 2001). For 5'-RACE rapid amplification of cDNA end (5'-RACE), total RNA (2 µg) extracted from S2 cells was reverse-transcribed into first-strand cDNA by using SuperScript III Reverse Transcriptase (Invitrogen) following the protocol supplied. The sequence of the gene-specific primer (GSP) utilised (P1; see position in Fig. 5) was: 5'-CCACTGACGTGGGCTCAA-3'. The removal of RNA complementary to the cDNA was performed with 2 U of *E. coli* RNase H (Invitrogen) by incubating the sample at 37 °C for 20 min. cDNA was then purified by using the NucleoSpin Extract II kit (Macherey-Nagel) and eluted in 40 µl of bidistilled water. A homopolymeric A-tail was added to the 3' end of purified cDNA (10 µl) using 30 U of recombinant Terminal Deoxynucleotidyl Transferase (Invitrogen) and 200 µM dATP through incubation at 37 °C for 10 min. The A-tailed cDNA was purified as above and eluted in 20 µl of bidistilled water. Purified A-tailed cDNA (6 µl) was amplified with an oligo(dT)-adaptor primer (5'-GACCACGCGTATCGATGTCGACTTTTCTTTTCTTTT-TTV-3') and the GSP 5'-TCAAGTTCGGTTTCTCTATAAATC-3' (P2; see position in Fig. 5). A nested PCR was then carried out using the nested GSP (P3: 5'-CCGTTAAGCATGAAGCATT-3') and the adaptor primer (5'-GACCACGCGTATCGATGTCGAC-3').

### 2.3. Cells culture and α-amanitin treatment

S2 cells were grown in Schneider's *Drosophila* Medium supplemented with 100 U/ml of penicillin, 100 µg/ml of streptomycin and 10% heat inactivated fetal calf serum (Cambrex) and seeded at a



$4 \times 10^6$ /ml concentration. After 24 h, 0.2  $\mu$ g/ml  $\alpha$ -amanitin (Sigma) was added to the culture medium, the cells were harvested 72 h after treatment and subjected to RNA extraction.

#### 2.4. TMG immunoprecipitation

Cells were washed in PBS and lysed in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM PMSF). After 15 min on ice, NP40 was added to final concentration 0.5%, nuclei were collected by centrifugation and lysed in RIPA buffer [(150 mM NaCl, 1% NP-40, 50 mM Tris, pH 7.5, 5 mM EDTA, 1 mM PMSF and protease inhibitor cocktail (Roche)]. Nuclear extracts were incubated with the K121 anti-2,2,7-trimethylguanosine Mouse Agarose Conjugate mAb (Calbiochem) for 2 h at 4 °C, followed by extensive washes with RIPA buffer. Finally, RNA was extracted from the supernatant and immunoprecipitated fractions and analysed by Real Time RT-PCR.

#### 2.5. FISH analysis

S2 cells were grown on gelatinized glass coverslips, washed briefly in phosphate-buffered saline (PBS: 100 mM  $\text{Na}_2\text{HPO}_4$ , 137 mM NaCl, 27 mM KCl pH 7.4) and fixed in PBS containing 4% paraformaldehyde, for 30 min on ice. The cells were rinsed twice with ice-cold PBSM (0.1%  $\text{MgCl}_2$ -PBS buffer) and permeabilized with PBST (0.1% Triton X-100-PBS buffer) for 60 s at room temperature, then rinsed twice with PBSM, equilibrated in pre-hybridization solution (50% formamide/2 $\times$  SSC) and hybridized for 3 h at 37 °C in 20  $\mu$ l of a mixture containing 0.02% BSA, 40  $\mu$ g of *E. coli* tRNA, 2 $\times$  SSC, 50% formamide, 20 ng of labelled probe. The oligonucleotide probes were: dU85 (Cy3-conjugated), A\*TAACGTCGTCACCATGACAAACAGCTTAGACCTAACTA; Dm46E3 (Fluorescein-conjugated), G\*TCTTTCCTCATAATCTCCGT-TAAGCCATGAAGCAG (Primm). After two rinses with pre-warmed wash solution (50% formamide/2 $\times$  SSC) at 37 °C for 20 min, cells were rinsed first with 2 $\times$  SSC and then with PBSM at room temperature for 10 min. Counter staining of the nuclei was performed by incubating the cells with the nuclear staining solution (0.5  $\mu$ g/ml DAPI in RNase free water) for 5 min at room temperature. Slides were mounted in MOWIOL 4–88 (Calbiochem).

#### 2.6. Whole mount embryo in situ hybridization

Whole mount embryo *in situ* hybridization, using single-stranded DIG-labelled probes obtained by PCR, and immunohistochemical staining of embryos were performed essentially as described previously (Riccardo et al., 2007).

### 3. Results

#### 3.1. The list of U79-related *Drosophila* snoRNAs

Modification of human 28S rRNA at the A3809 residue, guided by the U79 snoRNA (Smith and Steitz, 1998), is highly conserved in vertebrates, invertebrates and plants, but absent in yeast. When the SnoScan program was used to search for *Drosophila* U79 putative orthologues we detected, among significant hits, five snoRNAs that were later validated experimentally by RT-PCR and Northern analysis (Supplemental Fig. 1). The list of detected snoRNAs included – in addition to the three tandem copies of Me28S-A2634, a snoRNA gene previously molecularly identified by Huang et al. (2005) on chromosome arm 3L – two novel snoRNAs mapping, respectively, on chromosome arms 2R and 3R (see Fig. 1A). According to their position on the polytene map, these novel specimens were named Dm46E3 (GenBank Acc: EU924798) and Dm83E4-5 (GenBank Acc: EU924799), respectively. For each of these genes, the antisense element predicted to target A2634, the *Drosophila* 28S residue equivalent to human A3809, was placed upstream the internal D' box (Fig. 1A).

A case-by-case examination revealed that these genes showed variegated genomic organizations and expression strategies (Fig. 2). The highly related Me28S-A2634 isoforms (named a, b and c) mapped tightly clustered at the 62C polytene region, within a small intergenic region separating the gene encoding the ribosomal protein Rpl23A from CG7974, a so far uncharacterized *Drosophila* gene transcribed from the opposite strand. As previously reported, Me28S-A2634 are generated by introns of a long, polyadenylated non-coding host transcript named dUhg 7 (Huang et al., 2005). This polycistronic arrangement is conserved also in humans where, as mentioned above, U79 is similarly intron-encoded by a long ncRNA, named GAS5, which harbours 10 box C/D snoRNAs within its introns (Smith and Steitz, 1998). Remarkably, structure of a long EST subsequently annotated in the *Drosophila* databases (GenBank Acc: BP542227) revealed that dUhg 7 overlaps with Rpl23A and CG7974 genes at their 3' ends (Fig. 2A), thus showing a more complex arrangement than previously suspected. Moreover, we annotated Me28S-A2634 isoforms also in the course of a parallel search for *Drosophila* U30 orthologues. Since human U30 is responsible for methylation of A3804 (Kiss-Laszlo et al., 1996), a site mapping only 5 nt apart that targeted by U79, we looked in more detail the Me28S-A2634/28S base pairing properties. Indeed, the complementarity to 28S extended throughout Me28S-A2634 D' boxes exactly by 5 nt, generating a longer duplex of 17 bp. However, this extended tract of complementarity is placed upstream an alternative, highly variant D' box (underlined in Fig. 1B) suggesting that, instead of guiding site-specific methylation of the A2629 residue, it may support a chaperone activity on rRNA folding. Anyhow, presence of this long antisense element may establish an intermediate evolutive link between the U30 and U79 guiding functions.

The two newly identified snoRNA genes, Dm83E4-5 and Dm46E3, were dispersed in the genome and did not share significant conservation each other or to the Me28S-A2634 copies. Noticeably, Dm83E4-5 spanned the first exon/intron junction of CG10284 (Fig. 2B), a still uncharacterized *Drosophila* gene predicted to encode a protein with serine-type endopeptidase inhibitor activity. In contrast with the majority of snoRNAs, whose intronic localization permits a concordant expression with the host gene, its peculiar arrangement predicts that Dm83E4-5 expression would be alternative to that of CG10284, being potentially able to negatively regulate host gene activity. The other snoRNA gene, Dm46E3, was instead embedded, with opposite polarity, within the first intron of *eiger* (Fig. 2C), a fly regulatory gene encoding a Tumor Necrosis Factor-like ligand involved in the regulation of programmed cell death, differentiation and immune response (Igaki et al., 2002; Kaupilla et al., 2003; Schneider et al., 2007).

When phylogenetic conservation of the five snoRNAs was checked by multiple genome alignments at the UCSC Genome Browser (<http://genome.ucsc.edu>), the Me28S-A2634 copies exhibited the highest sequence conservation, with their presence tracked back to the *D. grimshawi* genome. Dm46E3 also displayed significant conservation within the *melanogaster* subgroup, whereas Dm83E4-5 was very poorly conserved, indicating that it was more recently evolved.

#### 3.2. Dm46E3 may target a developmentally regulated isoform of the spliceosomal U1 snRNA

Looking carefully at the predicted secondary structures of the duplexes between the *Drosophila* 28S rRNA and the guide antisense elements of each snoRNA (see Fig. 1A), we noticed that the duplex formed by Dm46E3 was atypical, being significantly shorter and weakened by two internal mismatches. Both features are expected to lower the efficiency at which the 28S rRNA may be targeted (Chen et al., 2007), prompting us to further investigate whether Dm46E3 may play additional functions. Since also spliceosomal RNAs (snRNAs) are known to represent common targets of snoRNA-directed modifications, we checked by BLAST search whether Dm46E3 could

## A

Assigned name	SNOSCAN prediction (score)	Genomic location Chromosome, config (start...end)	Predicted guide duplex between snoRNAs and 28S rRNA	snoRNA location within gene sequence(s)	snoRNA estimated length
Me28S-A2634a	27.91	3L,(1647591...1647499) NT_037436.1	5' AGUAGGAAUCUCAUUGA 3'       3' UCAUCCUUAGAGC 5' *	dUH7	73
Me28S-A2634b	28.77	3L,(1647372...1647296) NT_037436.1	5' AGUAGGAAUCUCAUUGA 3'       3' UCAUCCUUAGAGC 5' *	dUH7	81
Me28S-A2634c	31.20	3L,(1647149...1647068) NT_037436.1	5' AGUAGGAAUCUCAUUGA 3'       3' UCAUCCUUAGAGC 5' *	dUH7	86
Dm83E4-5	22.18	3R,(2201095...2201214) NT_033777.1	5' AGCAGGAAUCUCAGUGA 3'          3' UCAUCCUUAGAGC 5' *	CG10284 (presumed serine-type endopeptidase inhibitor) Exon 2/Intron 2	140
Dm46E3	21.10	2R,(51415530...5141430) NT_033778.1	5' GUACGAAUGUCAGUGA 3'           3' CAUCCUUAGAGC 5' *	CG12919 (egr, encodes TNF-like ligand) Intron 1 Opposite strand	137

## B

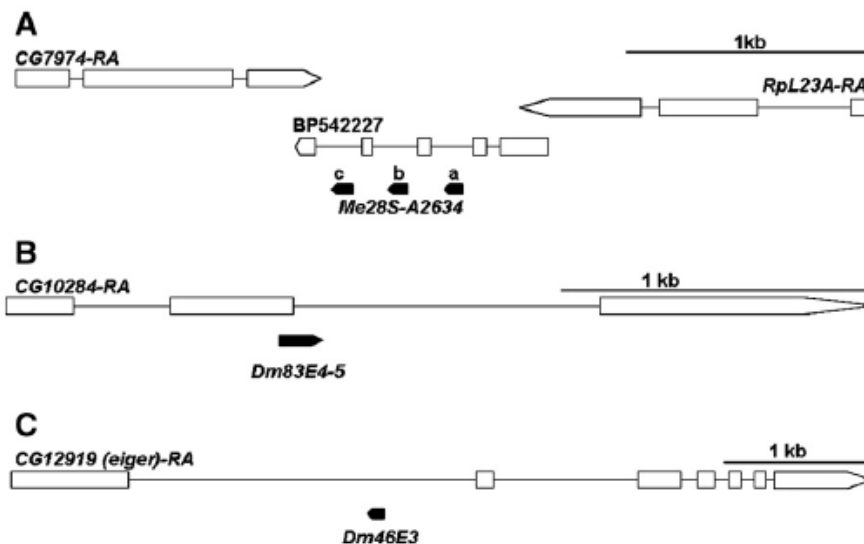
Me28S-A2634a	5' AGUAGGAAUCUCA <u>UUG</u> AAUAA 3'       3' UCAUCCUUAGAGCAAUUA 5' *	<b>Target sites of U30 and U79 on human 28S rRNA, and Me28S-A2634 on Drosophila 28S rRNA</b>  <b>Hs snoRNA:</b> U30 U79 <b>Residues:</b> 3804 3809  CGCGCATGAATGGATGA <u>ACGAGATT</u> CCCACTGTCCC Hs CGCGCATGAATGGATTA <u>ACGAGATT</u> CTCTACTGTCCC Dm  <b>Residues:</b> 2629 2634 <b>Dm snoRNA:</b> Me28S-A2634
Me28S-A2634b	5' AGUAGGAAUCUCA <u>UUG</u> GCUC 3'       3' UCAUCCUUAGAGCAAUUA 5' *	
Me28S-A2634c	5' AGUAGGAAUCUCA <u>UUG</u> AGCUC 3'       3' UCAUCCUUAGAGCAAUUA 5' *	

**Fig. 1.** (A) List of experimentally validated *Drosophila* U79 orthologues found by SnoScan analysis (Lowe and Eddy, 1999). In the predicted guide duplexes, C/D guide RNA sequences in a 5' to 3' orientation are shown in upper strands (D' box motifs underlined), while rRNA sequences in a 3' to 5' orientation are shown in lower strands. The positions of predicted methylation sites are indicated by an asterisk. Numbering of *Drosophila* 28S rRNA is according accession: M21017. Only Me28S-A2634 and Dm83E4-5 recognize rRNA target sequences by complementary antisense elements conforming to the canonical pairing constraints (Chen et al., 2007). (B) Complementarity of Me28S-A2634 isoforms to 28S extends throughout canonical D' boxes. On the left, alternative base-pairing between Me28S-A2634 isoforms and 28S rRNA. Canonical D' boxes, shown in A, are boxed, whereas alternative, highly degenerate D' motifs are underlined. The asterisk marks the position of the *Drosophila* 28S A2629 residue, equivalent to the A3804 site targeted by human U30 (on the right; numbering of human 28S rRNA is according accession: NR\_003287).

recognize any *Drosophila* snRNA sequence. Indeed, we observed that it can target efficiently U1:82Eb (U1b; see Fig. 3A), a developmentally regulated variant isoform of *Drosophila* U1 snRNA (Lo and Mount, 1990). Intriguingly, the same two nucleotide substitutions that interrupt its base pairing to 28S enabled Dm46E3 to match perfectly the U1b sequence (Fig. 3A), strongly supporting the view that these two point mutations changed the properties of this specimen and created a new guiding function. Acquisition of this novel functional role would have been facilitated by presence of multiple intact copies of Me28S-A2634 and Dm83E4-5 genes that preserved the 28S rRNA essential modification function, allowing the Dm46E3 guide element to evolve more freely and capture U1b as a new target. Among the seven U1 snRNA isoforms documented in *D. melanogaster*, U1b has unique structural and expression features. This variant isoform is in fact significantly longer than the prototype U1 sequence (designated as U1a), being characterized by presence of additional 5' and 3' tails,

as well as by a single nucleotide change (a G to T transversion at position 134; bold and underlined in Fig. 3B) within the central sequence common to all *Drosophila* U1 isoforms (Lo and Mount, 1990; Fig. 3B). U1b is also characterized by a specific expression profile. In fact, U1a predominates during later stages of the life cycle and in various terminally differentiated cell lines, whereas U1b is abundant during the embryonic stages, where it represents the major variant form in cell lines that possess the ability to differentiate (Lo and Mount, 1990). Dm46E3 can recognize U1b at its specific 3' tail, being potentially able to modify the terminal region that follows the SM binding site, at the A234 residue (marked by upper asterisk in Fig. 3B). Worth noting, the Dm46E3/U1b base-pairing may elicit functional consequences even in the absence of a site-specific modification, for example by influencing the U1b secondary structure and/or the spliceosomal assembly. To further address the Dm46E3 functional role, we then examined whether it localized in the Cajal body (CB), as





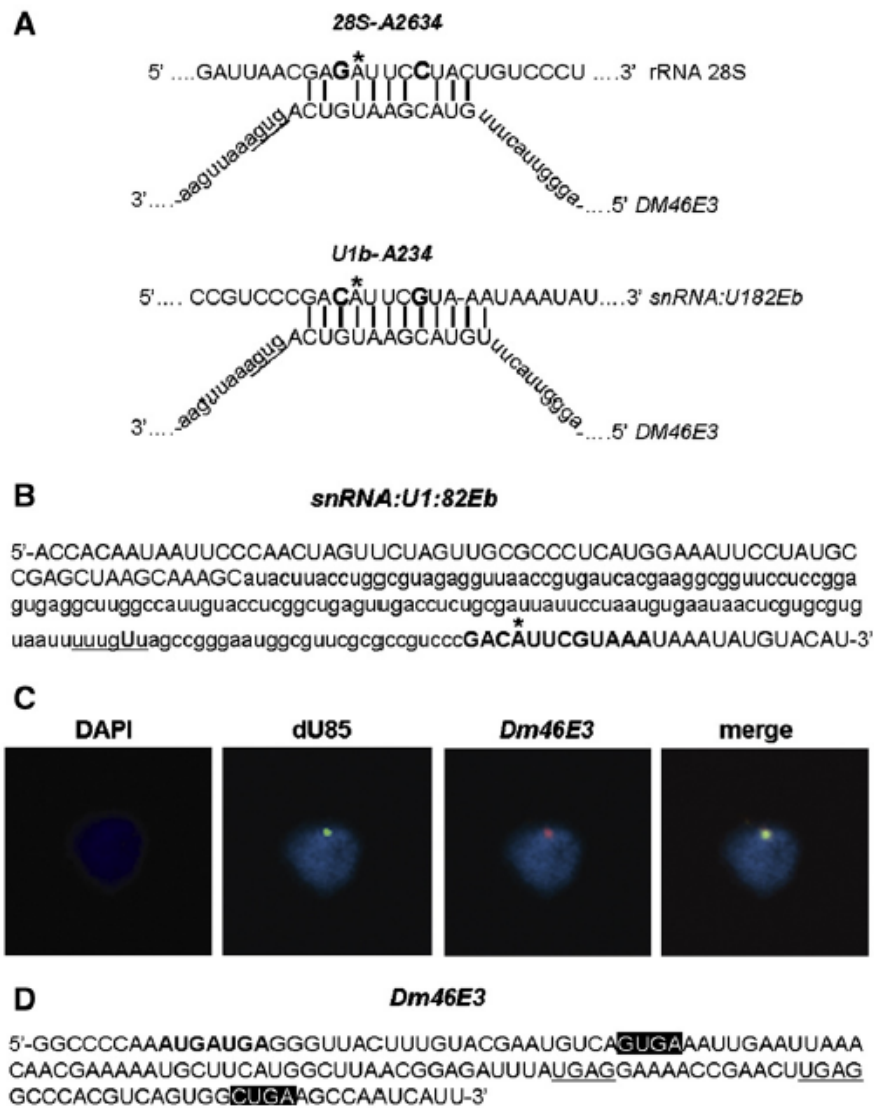
**Fig. 2.** Schematic diagram of *Me28S-A2634*, *Dm83E4-5* and *Dm46E3* genetic arrangement. In A, the structure of the longer EST (Accession: BP542227) representative of the dUhg 7 is outlined. No EST representative of the *Dm83E4-5* and *Dm46E3* transcripts was instead found in the databases.

predicted for snoRNAs able to guide modifications of the RNA Pol II-transcribed snRNAs, such as U1, U2, U4 and U5 (Richard et al., 2003). As determined by fluorescent *in situ* hybridization (FISH) of S2 cells, *Dm46E3* concentrates within one discrete dot-like nucleoplasmic domain (Fig. 3C). This region also concentrated dU85 (Richard et al., 2003; Liu et al., 2006), a *Drosophila* scaRNA known to be enriched in these nucleoplasmic organelles that we used as CB marker. As shown in Fig. 3D, *Dm46E3* sequence also contains Cajal body-specific localization signals, or CAB boxes (UGAG; Henras et al., 2004). Altogether, these features indicated that *Dm46E3* is indeed a novel *Drosophila* scaRNA, so hereafter we refer to it as *scaDm46E3*. Both localisation and base-pairing properties of this novel scaRNA strongly support the view that it may effectively be involved in U1b targeting, allowing to predict the first specific modification occurring on this developmentally regulated U1 isoform.

### 3.3. Expression strategy of *scaDm46E3*

Northern blot developmental analysis established that *scaDm46E3* was constitutively expressed along the *Drosophila* life cycle (see Supplemental Fig. 1C). Considering the intriguing antisense arrangement of this scaRNA, we attempted to investigate in more detail the molecular mechanisms underlying its expression. No EST with the same transcriptional orientation was annotated in the databases, suggesting that *scaDm46E3* may be independently transcribed as a singlet. Although rare, independent transcription of *Drosophila* snoRNAs has been reported to depend on either Pol II or Pol III machinery (Isogai et al., 2007), so we wished to address this point by checking *scaDm46E3* sensitivity to  $\alpha$ -amanitin. *Drosophila* cultured S2 cells were then treated for 72 h with 0.2  $\mu$ g/ml of  $\alpha$ -amanitin as described under Materials and methods, and the effectiveness of the treatment determined by real time RT-PCR. In these experiments, the level of three Pol II transcripts ( $\alpha$ Tub and *mfl* mRNAs, *snRNA U1*), together with that of the Pol III-transcribed 7SL RNA (Isogai et al., 2007), were followed as internal controls; all data were normalized against the level of the Pol I-dependent 5.8S rRNA, insensitive to  $\alpha$ -amanitin. As shown in Fig. 4B, at the conditions applied the treatment left totally unaffected the amount of the 7SL RNA, whereas resulted in an approximately 5-fold reduction of the levels of *scaDm46E3* and all Pol II-transcribed controls, clearly indicating that *scaDm46E3* expression was Pol II-dependent. Given that the majority of snRNAs and snoRNAs independently synthesized by Pol II has a cap structure

resembling that found on messenger RNAs, but with a characteristic trimethylation of the cap guanosine (Tycowski et al., 2004), we next checked whether *scaDm46E3* would possess a trimethylguanosine (TMG) cap. Nuclear RNA extracted from S2 cells was then treated with the monoclonal K121 anti-TMG antibody, and the relative abundance of *scaDm46E3* between the immunoprecipitated or supernatant fractions checked by quantitative real-time RT-PCR analysis. The level of the TMG capped U1 snRNA was followed as positive control, while that of the intron-encoded *DmSnR60* snoRNAs, hosted by the *mfl* gene (Riccardo et al., 2007), was taken as a negative reference. As shown in Fig. 3C, the level of U1 was more than 25-fold enriched by precipitation via the anti-TMG antibody, whereas *scaDm46E3* and the intronic *DmSnR60* isoforms failed to exhibit any enrichment, pointing out that only a trascurable, if any, fraction of *scaDm46E3* shows cap hypermethylation. Altogether, these data indicated that *scaDm46E3* is transcribed by Pol II most plausibly as a longer, perhaps rapidly processed, precursor. To better define this aspect, we performed 5' RACE experiments to precisely map the scaRNA 5' end. Noticeably, these experiments revealed that *scaDm46E3* transcripts exist as two species having distinct 5' ends mapping more than 400 bp apart (Fig. 5A–C). Nucleotide sequencing of the short fragment confirmed the 5' end expected for the *scaDm46E3* mature form, whereas that of the longer, less abundant species revealed presence of an unspliced 5' extension of about 410 nt. Presence of this unspliced extension was further confirmed by RT-PCR analysis, in which two upstream forward primers (P4/P5) were used in combination with the same reverse primer (P3) derived from the *scaDm46E3* sequence. As shown in Fig. 5B, all primer pairs yielded in fact positive amplification of a fragment with the expected size. Noticeably, the remote 5' end mapped in correspondence of a run of adenines, raising the possibility that the oligo(dT)-adaptor primer may have internally annealed to a longer transcript originating further upstream. Intriguingly, 5' processing has recently been reported for *snoRNA:314* and *snoRNA:644*, two *Drosophila* Pol III-transcribed C/D snoRNAs (Isogai et al., 2007). Although the functional significance of these 5' extensions was unclear, it is possible that annotation of several snoRNAs is likely to reflect the size of the mature forms and not the true transcriptional start site. Indeed, specific cleavage of several classes of ncRNAs has recently been reported, revealing distinct patterns of small RNA biogenesis so far ignored (Isogai et al., 2007; Kawaji et al., 2008; Ganesan and Rao, 2008) and suggesting that a part of small ncRNAs may be present in multiple longer forms, which may eventually play different biological roles.



**Fig. 3.** *Dm46E3* target specificity and intranuclear localization. (A) Base pairing of the *Dm46E3* D' antisense element with 28S (up) or snRNA:U:82Eb (U1b; down); the two bases that interrupt complementarity to 28S (marked in bold) allow perfect matching to U1b. (B) *Drosophila* U1b sequence; the U1b 5' and 3' tails specific tails (capital) flank the central prototype U1a sequence (lowercase) containing the SM binding site (underlined) which shows a single G to T transversion (bold capital); the tract of complementarity to the *Dm46E3* D' antisense element (shown in bold capital) contains the A234 target site (marked by the upper asterisk). (C) *in situ* localization of *Dm46E3* RNA in S2 cells. *Dm46E3* and dU85 RNAs were visualized by sequence-specific fluorescent oligonucleotide probes, as described in Materials and methods. The merged image shows that *Dm46E3* RNA co-localizes with the Cajal body marker dU85 RNA (Richard et al., 2003). (D) *Dm46E3* nucleotide sequence. The C box is shown in bold; D and D' sequences are boxed; two canonical CAB boxes are underlined.

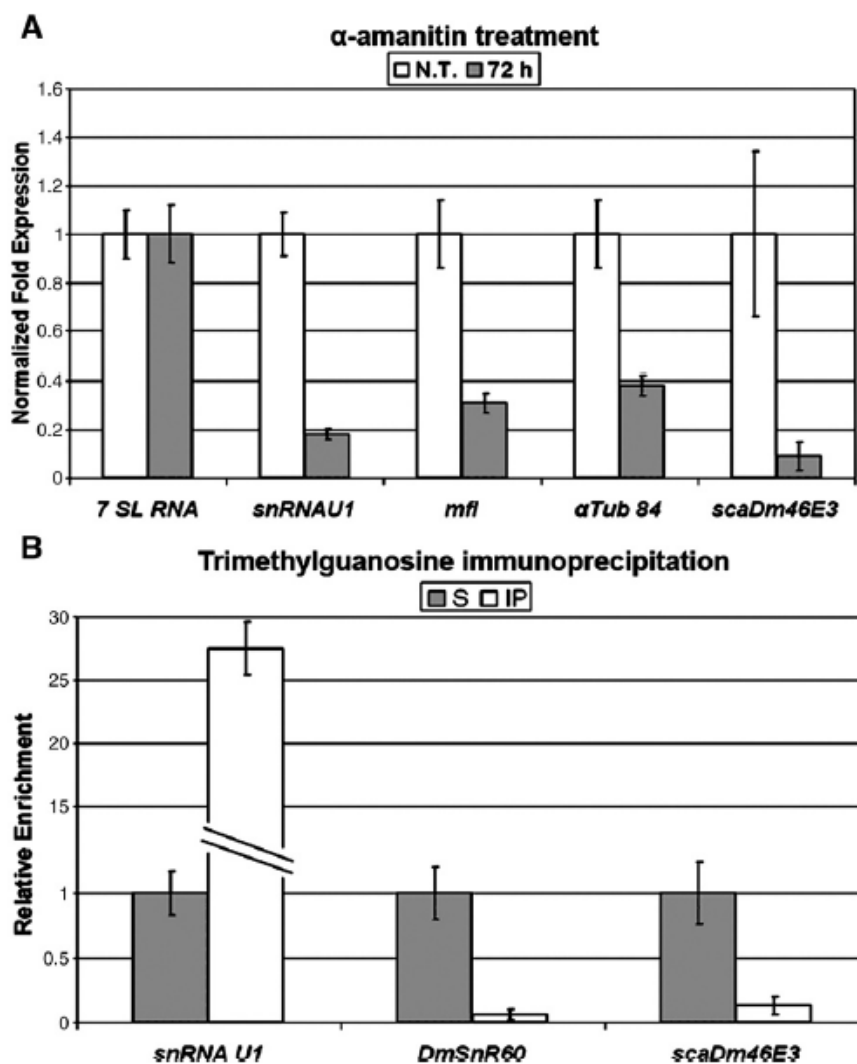
Indeed, the observation that *scaDm46E3* is embedded within a large portion of *egr* intronic sequence that is remarkably conserved across the *melanogaster* subgroup (Fig. 5D), further supporting the view that it might derive from a longer precursor.

#### 3.4. Mutual regulation of the *scaDm46E3/egr* sense-antisense gene pair

Antisense transcription is a widely exploited mechanism to regulate gene expression in eukaryotic cells (Lehner et al., 2002; Osato et al., 2007). Considering that, as demonstrated in many cases, an antisense RNA can inhibit accumulation of another mRNA to which is complementary, we attempted to investigate in more detail the regulatory mechanisms governing the expression of the *scaDm46E3/egr* gene pair. In order to assess whether the expression of the two sense-antisense genes could be mutually regulated, we checked *scaDm46E3* accumulation in the *egr*<sup>3</sup> mutant strain, where *egr* transcription was shown to be strongly reduced (Igaki et al., 2002). *egr*<sup>3</sup> flies carry a deletion that removes about 1.5 kb of *egr* genomic

sequences, but leaves the *scaDm46E3* coding, upstream and downstream surrounding sequences intact (Igaki et al., 2002; see also Fig. 6A). When the *scaDm46E3* expression level of wild-type and *egr*<sup>3</sup> flies was compared by quantitative real-time RT-PCR analysis, we found that it was more than 10-fold increased in the mutants (Fig. 6B), indicating that *egr* transcriptional efficiency can significantly influence the expression of its antisense snoRNA. Moreover, the finding that *scaDm46E3* accumulates at an even higher amount in S2 cultured cells (Fig. 6B) suggested that its expression may be up-regulated in proliferating cells, or eventually down-regulated in adult flies by regulatory factors not present in this cell line. Mutual regulation by sense-antisense transcription may make gene expression amenable not only to temporal, but also to spatial regulation, so that RNAs that form sense-antisense pairs frequently exhibit reciprocal expression patterns. In the case in point, *egr* is known to exhibit a specific expression pattern during embryogenesis and in larval tissues, being predominantly expressed in the nervous system (Igaki et al., 2002; Kaupilla et al., 2003). We thus asked whether *scaDm46E3* might





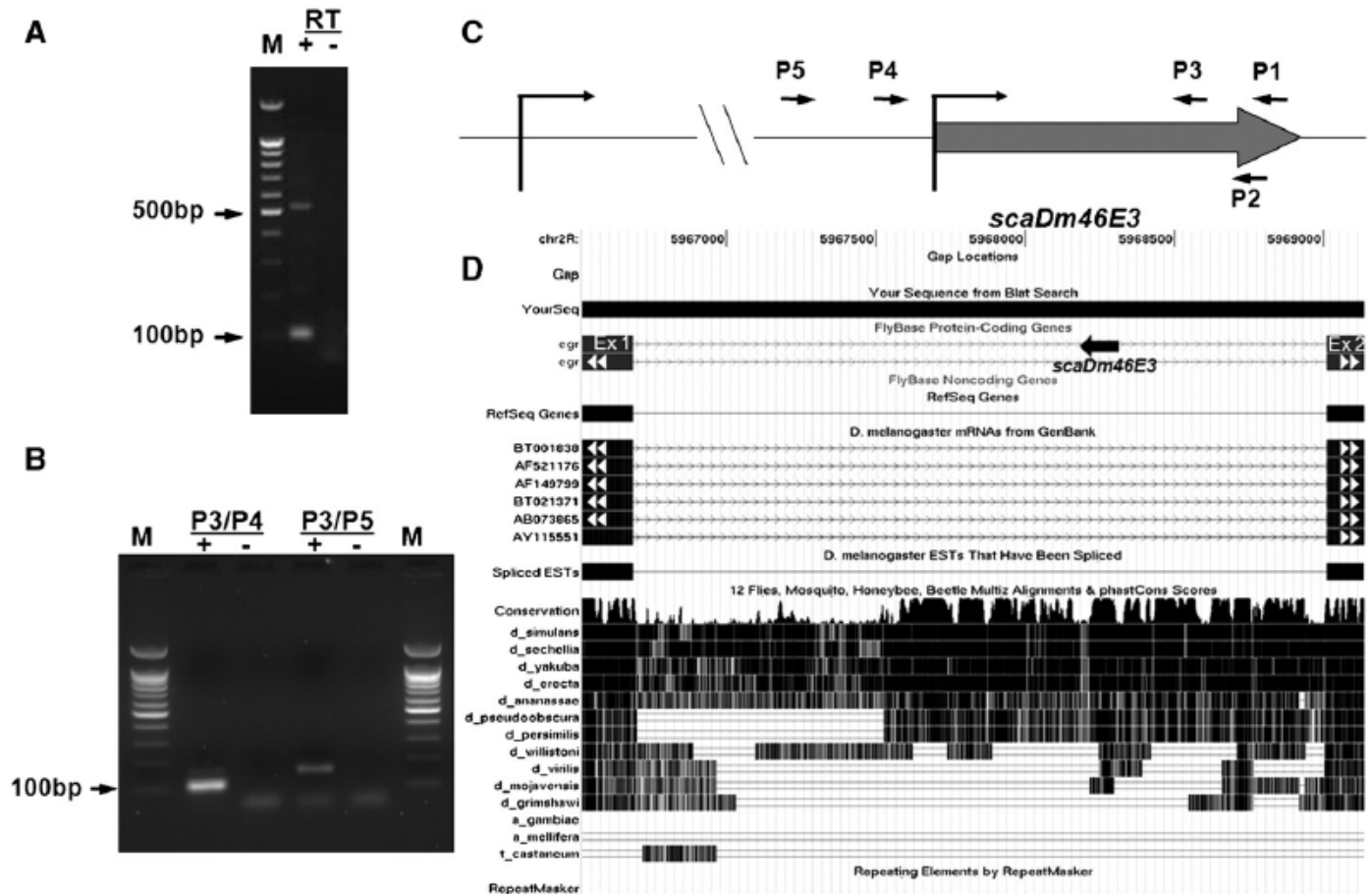
**Fig. 4.** *ScaDm46E3* is Pol-II-dependent and lacks Trimethylguanosine cap. (A) Sensitivity of *scaDm46E3* accumulation to  $\alpha$ -amanitin, showing its dependence on Pol II machinery. RNA from untreated or treated (see Materials and methods) cells was analysed by Real time RT-PCR. Data were normalized to 5.8 rRNA expression and presented relative to the untreated cells; they represent the average of at least three independent experiments. In treated cells, *scaDm46E3* expression decreases similarly to the examined Pol II-dependent transcripts (*αtub*, *mfl*, *snRNA U1*), whereas the Pol III-transcribed 7SL RNA was totally unaffected. (B) Quantitative real-time RT-PCR analysis on the immunoprecipitated (IP) or supernatant (S) fractions of nuclear RNA extracted from S2 cells after incubation with the K121 monoclonal anti-TMG antibody (Calbiochem). Trimethylguanosine capped U1 snRNA shows a marked IP-enrichment, whereas no IP-enrichment was observed for *scaDm46E3* or the *DmSnR60* negative control.

similarly display a spatially-restricted expression profile during embryogenesis. To this purpose, staged whole-mount embryo preparations were analysed by *in situ* hybridization with *scaDm46E3*-specific probes. As shown in Fig. 7A, a strong ubiquitous hybridization signal was detected in pre-blastoderm embryos (stages 1–3), indicating a high level of maternal contribution. Considering that U1b is maternally deposited in the developing oocytes (Lo and Mount, 1990), the high-level expression of *scaDm46E3* observed in early embryos nicely correlates with its predicted role on U1b modification. Immediately after, *scaDm46E3* expression concentrated at posterior pole, although it remained excluded from the pole cells, the first cells formed in the *Drosophila* embryo and the germ cell precursors (stage 4; Fig. 7B). During early gastrulation (stages 6–7; Fig. 7C) the signal marked the amnioproctodeal invagination, where the primordia of the hindgut and the posterior midgut are formed, and remained concentrated there also in the following stages. High-level expression of *scaDm46E3* marked in fact these tracts of the developing intestine also during late gastrulation and in segmented embryos (stages 9–10; Fig. 7D, E). At these stages, *scaDm46E3* also marks the neurogenic region, a territory in which its expression coincides with

that of *egr* (Igaki et al., 2002; Kauppila et al., 2003). Finally, in late segmented embryos *scaDm46E3* is also prominently upregulated in the amnioserosa (stage 14; Fig. 7F), a dorsal stretch of extraembryonic epithelium that connects the dorsal and ventral halves of the segmented germ band and is progressively covered by lateral ectodermal cells undergoing dorsal closure, the last morphogenetic movement during embryogenesis. An obvious conclusion from these *in situ* experiments is that the spatial expression of *scaDm46E3* is dynamically regulated during embryogenesis. Moreover, it concentrates in the developing intestine and in the amnioserosa, two embryonic regions not marked by *egr*, whose expression is restricted at the dorsal furrows until embryonic stage 9, and thereafter concentrates essentially in the developing nervous system (Igaki et al., 2002; Kauppila et al., 2003).

#### 4. Discussion

Although ribose 2'-O-methylation represents the most common nucleotide modification of rRNAs and snRNAs, a significant portion of C/D snoRNA genes still remains to be identified in most genomes.

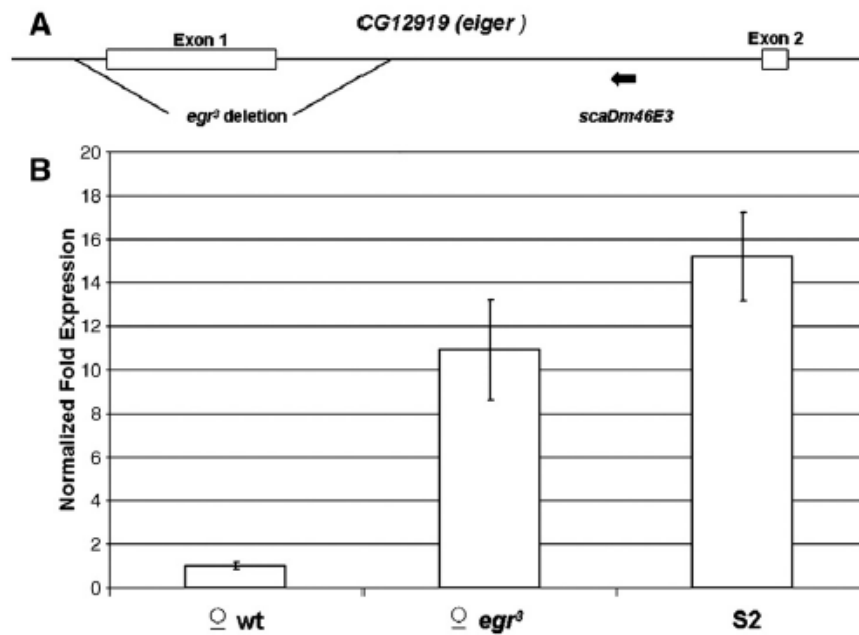


**Fig. 5.** *ScaDm46E3* transcripts exist as two different species differing at 5' end. (A) 5' RACE products obtained after nested PCR-amplification (performed with the P3/adaptor primer pair; see Materials and methods). For 5' RACE, retrotranscription makes use of the P1 gene-specific primer, and the P2/oligo(dT)-adaptor primer pair was used in the first PCR amplification (see (C) for position of primers). A 100 bp amplification product corresponding to the *scaDm46E3* mature form and a highest accessory band are detected in presence (+RT), but not in absence of reverse transcriptase (-RT), indicating absence of contaminating genomic DNA. Lane M: 100 bp Plus Molecular Weight Marker (Applchem). (B) RT-PCR analysis on S2 RNA confirming the presence of *scaDm46E3* long precursor form. Both P3/P4 and P3/P5 primer pairs show positive amplification of fragments with the expected size in presence (+), but not in absence of reverse transcriptase (-), indicating absence of contaminating genomic DNA. (C) Schematic representation of the primers used in A and B. (D) Blat search of *egr* exon1-intron1-exon2 sequence indicates a marked conservation of a large intron 1 portion across the *melanogaster* subgroup.

Indeed, identification of specimens showing atypical structure or genomic location, poor expression or unknown function is expected to be particularly elusive. Our results testify how a combination of bioinformatics, comparative genomics and molecular approaches may synergistically contribute to uncover the hidden layer of the snoRNA transcriptome, extending our knowledge about snoRNA genes, their genomic arrangement, expression strategies and evolutive mechanisms. In previous works, these combined approaches already proved to be successful, allowing us to identify the *Drosophila* counterparts of two out of ten GASS intron-encoded snoRNAs (Accardo et al., 2004; Riccardo et al., 2007). In addition to *Me28S-A2634* isoforms, which may represent ancient U79 orthologues, two novel RNAs, *snoDm83E4-5* and *scaDm46E3*, were identified in the course of our bioinformatics search. Duplication and dispersion in the genome, followed by genetic drift, or convergent evolution are possible mechanisms accounting for the origin of these dispersed genes. An additional mechanism has been recently evoked for the evolution of mammalian snoRNAs, for which retroposition of a parental copy, followed by genetic diversification, has been hypothesized (Weber, 2006; Luo and Li, 2007). Whatever process creates a new snoRNA copy and establishes genetic redundancy, over evolutionary time divergence dynamic mechanisms can increase snoRNA diversity to eventually give rise to an inactive isoform that would decay during the course of evolution or, alternatively, originate a new RNA-modification function that might be lineage or species specific.

To this regard, it is widely accepted that point mutations in motifs related to C, D or D' box sequences or in the antisense elements might have a strong impact on the functionality of a snoRNA copy, playing a key role on the evolution of novel functions. By this mechanism, snoRNAs may evolve new target site complementarities and the gain, loss and change of targets of over relatively short evolutionary times may allow these small ncRNAs to continuously contribute to the changing needs of cells and genomes. In this light, the *scaDm46E3* gene described nicely shows how subtle changes within the antisense element can drastically affect the target specificity, creating a new guide function. Two point mutations within the *scaDm46E3* antisense element were in fact sufficient to produce a target switch from 28S rRNA to the U1b snRNA. Although a more conclusive verification of U1b modification is necessary, it can be surmised that even in the absence of site-specific modification the *scaDm46E3*/U1b base-pairing may elicit functional consequences, possibly by influencing snRNP assembly. Intriguingly, out of the 63 fly C/D box snoRNAs identified so far, only five were expected to guide internal methylations of snRNAs, with U2, U5, and U6 snRNAs representing the predicted targets (Huang et al., 2005). Thus, the *scaDm46E3* antisense element leads to envisaging the first methylation on *Drosophila* U1 snRNAs, with the U1b embryonic variant expected to be specifically targeted. Intriguingly, the finding that *scaDm46E3* is expressed with a dynamic, spatially restricted expression profile during *Drosophila* embryogenesis raises the possibility



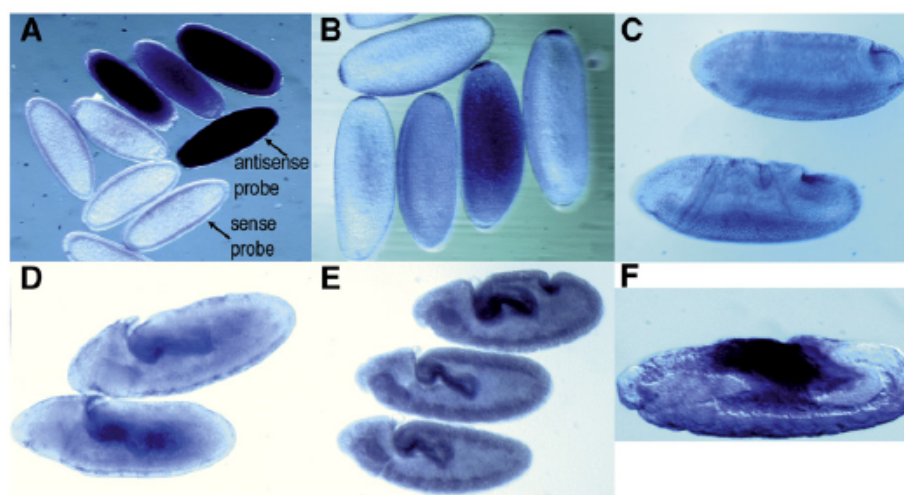


**Fig. 6.** *ScaDm46E3* is over expressed in *egr*<sup>3</sup> mutant flies and in S2 cells. (A) Schematic diagram of the 1.5 kb deletion carried by *egr*<sup>3</sup> flies and relative position of the *scaDm46E3* gene. (B) Amounts of *scaDm46E3* transcripts in wt and *egr*<sup>3</sup> adult flies and S2 cells, determined by real-time RT-PCR. Data represent the average of at least three independent experiments. The markedly higher *scaDm46E3* accumulation observed in the *egr*<sup>3</sup> strain indicates that *egr* transcriptional efficiency influences the expression of its antisense scaRNA. The even more pronounced *scaDm46E3* accumulation level observed in S2 cell suggests that this snoRNA is up-regulated in proliferating cells, or down regulated in adult flies.

that a subset of snRNA modifications might be developmentally regulated. According to this hypothesis, it seems conceivable that snoRNAs able to target spliceosomal RNAs may act as critical splicing modifiers, mediating in a subtle way a variety of biological effects in higher eukaryotes.

In the evolution of a snoRNA copy, the activity of a new gene often relies on the establishment of a novel expression strategy that, once established, may exert diverse regulatory effects on overlapping or surrounding genetic regions. Intriguingly, both the *snoDm83E4-5* and *scaDm46E3* genes show unusual genomic location and follow expression strategies totally different from that displayed by the majority of *Drosophila* C/D snoRNAs. While most C/D snoRNAs are generally intron-encoded by either protein-coding or non-coding

host genes (HGs), and released after splicing of HG's primary transcript, expression of *snoDm83E4-5* and *scaDm46E3* is instead envisaged to play a negative role on the activity of their overlapped genes. Based on its location at an exon–intron junction, *snoDm83E4-5* expression is in fact predicted not only to be splicing-independent, but even alternative to the production of its HG's spliced mRNA. Similarly, expression of the *scaDm46E3/egr* sense–antisense gene pair appears to be mutually regulated, making plausible that *scaDm46E3* expression may regulate *egr* activity via RNAi or by modulating splicing of *egr* pre-mRNA. A significant percentage of several genomes, including those of mammals, is known to be transcribed from both strands, and a number of well-characterized antisense transcripts appear to play regulatory roles in relation to



**Fig. 7.** *ScaDm46E3* shows a dynamic expression profile during *Drosophila* embryogenesis. *Drosophila* embryos were labelled with digoxigenin-labelled antisense probes at different developmental stages. In (A) and (B) pre-blastoderm embryos; the strong ubiquitous signal in (A) denotes maternal contribution. (C–F) lateral views of early gastrulating (C), segmented embryos (D, E) and late embryos undergoing dorsal closure (F). The signal concentrates from the primordia in the hindgut and the posterior midgut, and later in the amnioserosa, before these cells are covered by the dorsal closure.

their sense gene (reviewed by Lapidot and Pilpel, 2006). In addition to work by RNA-directed mechanism, symmetrical transcription may be critical in regulating the state of a large chromosomal domain around a gene, leading to alterations of chromatin structure, DNA methylation, as well as promoter exclusion or competition. Although further experiments are required to understand what mechanism is actually responsible for *Dm46E3/egr* mutual regulation, it is reasonable to assume that it plays a role in ensuring that these two genes maintain the distinct spatial expression profiles observed during embryogenesis. Considering that, as member of the tumor necrosis factor (TNF) family, *egr* plays an important role in the regulation of cellular proliferation, differentiation and apoptosis through the activation of the JNK pathway (Igaki et al., 2002), further characterization of *scaDm46E3* may be relevant to understanding the molecular mechanisms governing the expression of this fly regulatory gene. Indeed, several *Drosophila* snmRNAs displayed active expression when located in an antisense orientations (Yuan et al., 2003; Accardo et al., 2004; Huang et al., 2005) suggesting that the small ncRNA transcriptome might have broad implications for fly gene regulation.

### Acknowledgments

We thank M. Dionne for furnishing the *egr*<sup>3</sup> strain and for helpful discussion. This work was supported by funds from the Assessorato alla Ricerca Scientifica, Regione Campania (Legge 5, annualità 2006) to M.F.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2009.02.005.

### References

- Accardo, M.C., et al., 2004. A computational search for box C/D snoRNA genes in the *Drosophila melanogaster* genome. *Bioinformatics* 20, 3293–3301.
- Bachellerie, J.P., Cavaille, J., Huttenhofer, A., 2002. The expanding snoRNA world. *Biochimie* 84, 775–790.
- Chen, C.L., Perasso, R., Qu, L.H., Amar, L., 2007. Exploration of pairing constraints identifies a 9 base-pair core within box C/D snoRNA–rRNA duplexes. *J. Mol. Biol.* 369, 771–783.
- Dong, X.Y., et al., 2008. SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum. Mol. Genet.* 17, 1031–1042.
- Ganesan, G., Rao, S.M., 2008. A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse. *RNA* 14, 1399–1410.
- Henras, A.K., Dez, C., Henry, Y., 2004. RNA structure and function in C/D and H/ACA s (no)RNPs. *Curr. Opin. Struct. Biol.* 14, 335–343.
- Huang, Z.P., Zhou, H., He, H.L., Chen, C.L., Liang, D., Qu, L.H., 2005. Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA* 11, 1303–1316.
- Huang, Z.P., Chen, C.J., Zhou, H., Li, B.B., Qu, L.H., 2007. A combined computational and experimental analysis of two families of snoRNA genes from *Caenorhabditis elegans*, revealing the expression and evolution pattern of snoRNAs in nematodes. *Genomics* 89, 490–501.
- Igaki, T., et al., 2002. Eiger, a TNF superfamily ligand that triggers the *Drosophila* JNK pathway. *EMBO J.* 21, 3009–3018.
- Isogai, Y., Takada, S., Tjian, R., Keles, S., 2007. Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. *EMBO J.* 26, 79–89.
- Kaupilla, S., et al., 2003. Eiger and its receptor, Wengen, comprise a TNF-like system in *Drosophila*. *Oncogene* 22, 4860–4867.
- Kawaji, H., et al., 2008. Hidden layers of human small RNAs. *BMC Genomics* 9, 157.
- Kishore, S., Stamm, S., 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311, 230–232.
- Kiss, T., 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* 109, 145–148.
- Kiss, T., Fayet, E., Jady, B.E., Richard, P., Weber, M., 2006. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb. Symp. Quant. Biol.* 71, 407–417.
- Kiss-Laszo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M., Kiss, T., 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85, 1077–1088.
- Lapidot, M., Pilpel, Y., 2006. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.* 7, 1216–1222.
- Lehner, B., Williams, G., Campbell, R.D., Sanderson, C.M., 2002. Antisense transcripts in the human genome. *Trends Genet.* 18, 63–65.
- Liu, J.L., Murphy, C., Buszczak, M., Clatterbuck, S., Goodman, R., Gall, J.G., 2006. The *Drosophila melanogaster* Cajal body. *J. Cell Biol.* 172, 875–884.
- Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>−(ΔΔC<sub>T</sub>)</sup> method. *Methods* 25, 402–408.
- Lo, P.C., Mount, S.M., 1990. *Drosophila melanogaster* genes for U1 snRNA variants and their expression during development. *Nucleic Acids Res.* 18, 6971–6979.
- Lowe, T.M., Eddy, S.R., 1999. A computational screen for methylation guide snoRNAs in yeast. *Science* 283, 1168–1171.
- Luo, Y., Li, S., 2007. Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs. *Nucleic Acids Res.* 35, 559–571.
- Mattick, J.S., 2004. The hidden genetic program of complex organisms. *Sci. Am.* 291, 60–67.
- Mattick, J.S., Makunin, I.V., 2006. Non-coding RNA. *Hum. Mol. Genet.* R17–R29 15 Spec. No. 1.
- Mourtada-Maarabouni, M., Hedge, V.L., Kirkham, L., Farzaneh, F., Williams, G.T., 2008. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J. Cell. Sci.* 121, 939–946.
- Mourtada-Maarabouni, M., Pickard, M.R., Hedge, V.L., Farzaneh, F., Williams, G.T., 2009. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 28, 195–208.
- Nakamura, Y., et al., 2008. The GAS5 (growth arrest-specific transcript 5) gene fuses to BCL6 as a result of t(1;3)(q25;q27) in a patient with B-cell lymphoma. *Cancer Genet. Cytogenet.* 182, 144–149.
- Osato, N., Suzuki, Y., Ikeo, K., Gojobori, T., 2007. Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics* 176, 1299–1306.
- Pontes, O., Pikaard, C.S., 2008. siRNA and miRNA processing: new functions for Cajal bodies. *Curr. Opin. Genet. Dev.* 18, 197–203.
- Riccardo, S., Tortoriello, G., Giordano, E., Turano, M., Furia, M., 2007. The coding/non-coding overlapping architecture of the gene encoding the *Drosophila* pseudouridine synthase. *BMC Mol. Biol.* 8, 15.
- Richard, P., Darzacq, X., Bertrand, E., Jady, B.E., Verheggen, C., Kiss, T., 2003. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* 22, 4283–4293.
- Rogelj, B., 2006. Brain-specific small nucleolar RNAs. *J. Mol. Neurosci.* 28, 103–109.
- Sambrook, J., Russell, D.W., 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Schneider, D.S., et al., 2007. *Drosophila* eiger mutants are sensitive to extracellular pathogens. *PLoS Pathog.* 3, e41.
- Smith, C.M., Steitz, J.A., 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell. Biol.* 18, 6897–6909.
- Stanek, D., et al., 2008. Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. *Mol. Biol. Cell* 19, 2534–2543.
- Tycowski, K.T., Aab, A., Steitz, J.A., 2004. Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr. Biol.* 14, 1985–1995.
- Vitali, P., et al., 2005. ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.* 169, 745–753.
- Weber, M.J., 2006. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet.* 2, e205.
- Yuan, G., Klambt, C., Bachellerie, J.P., Brosius, J., Huttenhofer, A., 2003. RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.* 31, 2495–2507.